1967

Preliminary test procedures and Bayesian procedures for pooling correlated data

Donna Jean Brogan Ruhl
Iowa State University

RUHL, Donna Jean Brogan, 1939-
PRELIMINARY TEST PROCEDURES AND BAYESIAN PROCEDURES FOR POOLING CORRELATED DATA.

Iowa State University, Ph.D., 1967
Statistics

PRELIMINARY TEST PROCEDURES AND BAYESIAN PROCEDURES FOR POOLING CORRELATED DATA

by

Donna Jean Brogan Ruhl

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1967

TABLE OF CONTENTS

DEDICATION

dedicated

to

jeffrey

february 23, 1965

february 4, 1967

I. INTRODUCTION

Many times when it is desired to estimate a parameter, observations on two or more different random variables are available. For example, if it is desired to estimate the incidence rate of some disease on a defined population, the two random variables could be past medical histories and current physical examinations, where the medical histories and current physical examinations may or may not be on the same subjects. Let μ_x and μ_y be the mean incidence rate for the medical history subjects and the physical examination subjects, respectively. If $\mu_x = \mu_y$, then it seems logical to pool the observations with appropriate weights to obtain one estimate of the incidence rate. If $\mu_x \neq \mu_y$, a simple procedure would be to use only the data on the parameter of primary interest, most likely μ_y in this example. Thus, it is necessary to know whether the two means are equal or not. Assuming that the answer to this question is not known a priori, a statistical test of significance may be made of the null hypothesis H_o: $\mu_y = \mu_x$ versus the alternative hypothesis H_A: $\mu_y \neq \mu_x$. If H_o is accepted, the pooled estimator is used; if H_o is rejected, the unpooled estimator is used. This procedure is known as a preliminary test of significance and subsequent estimation or, according to Bancroft (1964), analysis and inference for an incompletely specified model using a preliminary test of significance. The model which describes the data at hand is incompletely specified because it is not known whether both the parameters μ_y and u_x are needed in the model (if $\mu_y \neq \mu_x$) or only one parameter μ is needed in the model (if $\mu_y = \mu_x = \mu$). Of course, there may be other parameters in the model besides μ_y and/or μ_x, but there is no uncertainty about the existence of the other parameters in the model.

There are also other kinds of estimation situations where observations are made on two or more random variables--for example, cases where ratio or regression estimators are appropriate. Concomitant information can also be used in structuring strata for stratified sampling or in probability proportional to size sampling.

Note that the preliminary test procedure of estimation has the experimenter behave as though the null hypothesis is true if it is not rejected. This is contrary to the emphasis in some applied statistics texts that failure to reject the null hypothesis does not imply that it is true. Many textbooks infer that failure to reject the null hypothesis can be attributed to a deficient sample size and/or lack of power rather than the veracity of the null hypothesis. However, Berkson (1942) disagrees with the philosophy that a low probability of obtaining the sample at hand disproves the null hypothesis and a <u>not</u> so low probability does <u>not</u> disprove the null hypothesis. He says that larger probabilities should be taken as prima facie evidence in favor of the null hypothesis and also points out that most experimentalists are typically engaged in finding evidence for affirmative conclusions rather than disproving hypotheses. Thus, in procedures involving preliminary tests of significance, the null hypothesis is not a "straw man" hypothesis to be knocked down, but a hypothesis which can be equally accepted or rejected.

The first paper to discuss preliminary tests and subsequent estimation was by Bancroft (1944), and in the following years papers appeared which discussed various aspects of estimation and/or testing of hypotheses after a preliminary test of significance. For the most part these papers described what practicing statisticians had actually been doing anyway, without benefit of the mathematical description of the bias, mean square error, and cumulative distribution function of their estimators and test statistics. Such descriptions, however, aided the practicing statistician in selecting the size (α level) of the preliminary test and in assessing the utility of such an estimation or testing procedure.

All of the papers discussed the preliminary test problem assuming that observations on the two random variables are already available, and the statistician has to decide whether to use all of the available observations or to use only the observations on the parameter of interest. If the auxiliary observations are not used in a pooled estimator, then the cost of collecting them is wasted to a large extent. Of course, these observations contribute to the conclusion that all the data do not estimate the same parameter, but the emphasis in most cases is on the resultant estimator or test of hypothesis. This thesis investigates the possibility of allocating observations between the variable of interest and the concomitant variable so as to minimize mean square error for a fixed cost or, alternatively, to minimize cost for a fixed mean square error. This gives an element of survey planning to the problem rather than the approach by other authors of having the data already collected and starting at that point.

The optimal sample allocation is quite likely to give unequal sample sizes for the two random variables $\tilde{x}$ and $\tilde{y}$. Most previous authors in this field have considered the case of two independent variables, where the sample sizes of the two variables have been either equal or unequal. Only one author (Kitagawa, 1963) has considered two random variables which are correlated; he discussed a sample of size n from a bivariate normal distribution. This thesis is primarily concerned with correlated samples, not necessarily of the same sample size. This requires selecting a bivariate random sample of size $n \geq 0$ from the bivariate distribution of $\tilde{x}$ and $\tilde{y}$ and then selecting additional independent observations on the x and/or y variables. Also considered in this thesis are some two-stage sampling schemes, where a decision about which variable to further sample is decided after the preliminary test is made.

Many authors are not satisfied with the classical approach to testing hypotheses and estimation, which is the approach used in the preliminary test schemes. Edwards _et al_. (1963) criticized the extreme tendency of classical tests to reject null hypotheses, and they advocated Bayesian methods which can either strengthen or weaken null hypotheses. Aitchison (1964) and Thatcher (1964) advocated tolerance regions or confidence intervals for parameters via the Bayesian approach. They emphasized that tolerance regions by the classical or frequentist approach have an error frequency in the long run, i.e., over repetition of the same experiment a large number of times. The meaning of this for an individual experimenter, doing an experiment just one time, is hard to specify. If tolerance regions are obtained from the Bayesian point of view, however, the error rate is for that particular, individual experiment, and this error rate is much more meaningful to the experimenter. Authors such as Lindley (1961), Aitchison (1964), and Savage (1961a) have pointed out that Bayesian methods, by always considering the posterior distribution of the parameters rather than the distribution of the statistics, avoid the distributional problems inherent in the frequentist approach. An increase in mathematical tractability can also be obtained by using a Bayesian approach when a conjugate prior or non-informative prior distribution is used on the parameter(s). Of course, there is a price to be paid for this greater tractability, and it is the choice of a prior distribution on the unknown parameters. Raiffa and Schlaifer (1961) give conjugate prior distributions for several data distributions such as normal, uniform, Poisson, binomial, etc. They also give some very practical advice about how to choose a prior distribution, along with a strong argument for the use of the Bayesian method. Lindley (1961) and Welch (1964), among others, have noted that the exact form of the prior distribution is irrelevant in large samples, since the maximum likelihood estimator of the posterior distribution of the parameters differs from the maximum likelihood estimator of the sample likelihood function only by a factor of $1/n$. Of course, this is true only if the prior distribution is fairly smooth and does not consist of, for example, a mass of probability one at the point c. Thus, it is only in small samples that the choice of the prior distribution gives way to a discrepancy between the two approaches.

The Bayesian approach in this dissertation consists of assigning three different prior distributions to the random variables $\tilde{\mu}_y$ and $\tilde{\mu}_x$ and then using the mean of the posterior distribution of $\tilde{\mu}_y$ as the estimator of μ_y. For two of the three prior distributions that are considered, the Bayesian approach results in greater mathematical tractability than the preliminary test approach. The resultant posterior mean always assigns a non-zero weight to the concomitant information, whereas the preliminary test approach assigns a weight of zero to the concomitant information whenever H_A: $\mu_y \neq \mu_x$ is accepted. Thus, the Bayesian approach seems to make more efficient use of the concomitant information.

In the extension of the general pooling problem to more than two random variables or populations, the Bayesian approach seems to have an advantage over the preliminary test approach. Suppose an estimator of μ_z is desired, where the three random variables $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ are trivariate normal with respective sample means $\bar{x}$, $\bar{y}$, and $\bar{z}$. There exist standard procedures to test the null hypothesis H_o: $\mu_x = \mu_y = \mu_z$ versus the alternative hypothesis H_A: H_o not true. If H_o is accepted, then a pooled estimator $(w_x\bar{x} + w_y\bar{y} + w_z\bar{z})$ is used, where w_x, w_y, and w_z are some appropriate weights. If H_o is rejected, then the simplest procedure is to use $\bar{z}$ as the estimator of μ_z. However, if this procedure is used, some useful data may be ignored. For example, a pooled estimator is still appropriate if H_A is of the form $\mu_y \neq \mu_x = \mu_z$ (pool $\bar{x}$ and $\bar{z}$) or $\mu_x \neq \mu_y = \mu_z$ (pool $\bar{y}$ and $\bar{z}$). To obtain such a specification of the complex alternative hypothesis requires comparisons of paired means. This paired comparison testing after the initial preliminary test of significance further complicates the derivation of the bias and mean square error. In general, this approach does not look too promising.

In the Bayesian approach for three random variables, however, the only complication required over the approach for the two random variables $\tilde{x}$ and $\tilde{y}$ is the additional specification of a prior distribution on $\tilde{\mu}_z$. It appears that the Bayesian approach to pooling means will extend to k random variables much more easily than the preliminary test approach.

An attempt to compare the Bayesian and classical approaches to the pooling problem is made. As Bartholomew (1964, p. 201) says:

> "I hope it is a sign of the times that papers have begun to appear which compare Bayesian and frequentist approaches to inference in the context of particular problems...."

II. LITERATURE REVIEW

The first paper to discuss the use of a preliminary test of significance prior to the estimation of a parameter was by Bancroft (1944). He discussed two problems--the estimation of the variance of a normal population and the estimation of a regression coefficient of an orthogonal polynomial equation. In the first problem a preliminary test of H_o: $\sigma_1^2 = \sigma_2^2$ versus H_A: $\sigma_2^2 < \sigma_1^2$ is made to determine whether or not the sample variances s_1^2 and s_2^2 should be pooled to form an estimator of σ_1^2. In the regression problem a preliminary test is made to determine whether the regression equation should contain one or two dependent variables, and then the regression coefficient is estimated on the basis of the conclusion drawn from the preliminary test. Bancroft obtained the bias and mean square error for these two estimation procedures and pointed out that, in general, estimators that result from preliminary tests are biased.

The first paper to discuss the problem of pooling means was by Mosteller (1948). He took two approaches to the problem--the classical, preliminary test approach and the Bayesian approach. Mosteller considered two independent sample means $\bar{x}$ and $\bar{y}$ from normal populations with means μ_x and μ_y, respectively, and common variance σ^2, denoted by $N(\mu_x, \sigma^2)$ and $N(\mu_y, \sigma^2)$. σ^2 is known, and each mean is based on n observations. It is desired to estimate μ_y. In the classical preliminary test approach Mosteller computed the test statistic

$$z = (\bar{y} - \bar{x})(\sigma \sqrt{2} \,/\, \sqrt{n}\,)^{-1} \tag{2.1}$$

and defined his estimator $\hat{\mu}_{yc}$ as

$$\hat{\mu}_{yc} = \begin{cases} (\bar{y} + \bar{x})/2 & \text{if } |z| < \xi_\alpha \\ \bar{y} & \text{if } |z| \geq \xi_\alpha \end{cases} \tag{2.2}$$

where ξ_α is the critical value from the N(0,1) distribution such that

$$\int_{-\xi_\alpha}^{\xi_\alpha} (2\pi)^{-\frac{1}{2}} e^{-t^2/2} \, dt = 1 - \alpha. \tag{2.3}$$

He computed the bias and mean square error (MSE) of $\hat{\mu}_{yc}$, where

$$\text{Bias } (\tilde{\hat{\mu}}_{yc}) = \tilde{E}(\hat{\mu}_{yc}) - \mu_y \tag{2.4}$$

and

$$\text{MSE } (\tilde{\hat{\mu}}_{yc}) = \tilde{E}(\hat{\mu}_{yc} - \mu_y)^2. \tag{2.5}$$

The bias and mean square error are functions of σ^2, n, α, and Δ, where Δ is the difference between the two population means, i.e.

$$\Delta = \mu_y - \mu_x. \tag{2.6}$$

Mosteller compared $\text{MSE}(\tilde{\hat{\mu}}_{yc})$ to the mean square error obtained when $\bar{y}$ is used all the time, i.e. $\text{MSE}(\tilde{\bar{y}}) = \sigma^2/n$, by investigating the disadvantage coefficient $C = \text{MSE}(\tilde{\hat{\mu}}_{yc})/\text{MSE}(\tilde{\bar{y}})$. Letting

$$\delta = \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}, \tag{2.7}$$

which is the difference between the two parameter means measured in terms of the standard deviation of the difference of the two sample means, Mosteller graphed C as a function of δ for various values of α. Providing δ is small, he stated that pooling is more advantageous the higher the significance level of the preliminary test (i.e. the lower α is). Pooling is disadvantageous for intermediate values of δ, and the disadvantage increases with decreasing α. As δ gets very large, the disadvantage coefficient approaches C = 1 asymptotically from above.

In the Bayesian approach, Mosteller considered $\tilde{\Delta} = \tilde{\mu}_y - \tilde{\mu}_x$ to be a normally distributed random variable with mean 0 and variance $a^2\sigma^2$. Although not stated explicitly in his paper, he also considered $\tilde{\mu}_x$ to have a fairly constant distribution over the range of the most likely sample results, which, in effect, is what Savage (1962) calls "precise measurement." He took as his estimator $\hat{\mu}_{yb}$ the value of μ_y which maximizes the posterior distribution of $\tilde{\mu}_y$, given $\bar{x}$ and $\bar{y}$, i.e.

$$\hat{\mu}_{yb} = \left[\bar{y} + \frac{\bar{x}}{1+na^2}\right]\left[1 + \frac{1}{(1+na^2)}\right]^{-1} . \tag{2.8}$$

Note that as $a^2 \longrightarrow 0$, $\hat{\mu}_{yb} \longrightarrow (\bar{x} + \bar{y})/2$; and as $a^2 \longrightarrow \infty$, $\hat{\mu}_{yb} \longrightarrow \bar{y}$, where $(\bar{x} + \bar{y})/2$ and $\bar{y}$ are the two possible estimators in the classical, preliminary test approach. This illustrates the fact that, in general, the Bayesian estimators tend to reduce to the classical estimators under certain limiting conditions. Mosteller also obtained the mean square error of $\tilde{\hat{\mu}}_{yb}$ by considering $E(\tilde{\hat{\mu}}_{yb} - \tilde{\mu}_y)^2$ with Δ fixed and then taking expectation over $\tilde{\Delta}$, where $\tilde{\Delta}$ is distributed $N(0, a^2\sigma^2)$. This yielded

$$\mathrm{MSE}(\tilde{\hat{\mu}}_{yb}) = \frac{\sigma^2}{n}\left[a^2 + \frac{1}{n}\right]\left[a^2 + \frac{2}{n}\right]^{-1} . \tag{2.9}$$

Note that as $a^2 \longrightarrow 0$, $\mathrm{MSE}(\tilde{\hat{\mu}}_{yb}) \longrightarrow \sigma^2/2n$; and as $a^2 \longrightarrow \infty$, $\mathrm{MSE}(\tilde{\hat{\mu}}_{yb}) \longrightarrow$

σ^2/n, which are the variances for the always pool procedure and the never pool procedure, respectively. Also, note that MSE($\tilde{\hat{\mu}}_{yb}$) < MSE($\tilde{\bar{y}}$) for all values of a^2, whereas $\tilde{\hat{\mu}}_{yc}$ does not have this property. Mosteller (1948, p. 241) remarked that "...this result is rather important because it backs up our intuitive feeling that we should use all the information at hand in making our estimates."

Bennett (1952) also considered the problem of pooling means and extended Mosteller's (1948) work. He first considered two independent populations $N(\mu_y, \sigma^2)$ and $N(\mu_x, \sigma^2)$ with σ^2 known, from which independent sample means $\bar{y}$ and $\bar{x}$, based on n_y and n_x observations respectively, are selected. He defined his test statistic to be

$$z_1 = (\bar{y} - \bar{x}) \left(\sigma\sqrt{n_y^{-1} + n_x^{-1}}\right)^{-1} \tag{2.10}$$

and defined his estimator μ_{y1} as

$$\mu_{y1} = \begin{cases} (n_y\bar{y} + n_x\bar{x})\,(n_y + n_x)^{-1} & \text{if } |z_1| < \xi_\alpha \\ \bar{y} & \text{if } |z_1| \geq \xi_\alpha \end{cases} \tag{2.11}$$

where ξ_α is defined in equation (2.3). He derived the cumulative distribution function and frequency function of $\tilde{\hat{\mu}}_{y1}$, and then calculated E($\tilde{\hat{\mu}}_{y1}$) and MSE($\tilde{\hat{\mu}}_{y1}$). Letting $\mu_y = 0$ and considering various values of n_x, n_y, and α, he computed the value of μ_x which maximizes Bias($\tilde{\hat{\mu}}_{y1}$). For moderate levels of significance (α = .20, .10, .05) the values of μ_x which maximize Bias ($\tilde{\hat{\mu}}_{y1}$) are approximately independent of α, whereas for high significance levels ($\alpha = .01, .001$) these values of μ_x become very large and the bias has a larger maximum.

In the second case Bennett considered two independent distributions $N(\mu_y, \sigma_y^2)$ and $N(\mu_x, \sigma_x^2)$, where $\sigma_x^2 \neq \sigma_y^2$ and both σ_y^2 and σ_x^2 are known. Independent sample means $\bar{y}$ and $\bar{x}$ are based on n_y and n_x observations respectively. He defined his preliminary test statistic as

$$z_2 = (\bar{y} - \bar{x})\ (\sigma_y^2 n_y^{-1} + \sigma_x^2 n_x^{-1})^{-\frac{1}{2}} \tag{2.12}$$

and his estimator as

$$\hat{\mu}_{y2} = \begin{cases} \left[\dfrac{n_y \bar{y}}{\sigma_y^2} + \dfrac{n_x \bar{x}}{\sigma_x^2} \right] \left[\dfrac{n_y}{\sigma_y^2} + \dfrac{n_x}{\sigma_x^2} \right]^{-1} & \text{if } |z_2| < \xi_\alpha \\ \bar{y} \quad \text{if } |z_2| \geq \xi_\alpha \end{cases} \tag{2.13}$$

He obtained the cumulative distribution function and frequency function of $\tilde{\hat{\mu}}_{y2}$ and then calculated $E(\tilde{\hat{\mu}}_{y2})$ and $MSE(\tilde{\hat{\mu}}_{y2})$.

In the third case Bennett considered two independent populations $N(\mu_y, \sigma^2)$ and $N(\mu_x, \sigma^2)$ where σ^2 is unknown. The independent sample means $\bar{y}$ and $\bar{x}$ are based on n_y and n_x observations, respectively, and the pooled sample variance is

$$s^2 = (n_x + n_y - 2)^{-1} \left[\sum_{i=1}^{n_y} (y_i - \bar{y})^2 + \sum_{j=1}^{n_x} (x_j - \bar{x})^2 \right]. \tag{2.14}$$

He defined his preliminary test statistic as

$$t = (\bar{y} - \bar{x}) \left(s\sqrt{n_y^{-1} + n_x^{-1}} \right)^{-1} \tag{2.15}$$

and his estimator $\hat{\mu}_{y3}$ as

$$\hat{\mu}_{y3} = \begin{cases} (n_y\bar{y} + n_x\bar{x})\,(n_y + n_x)^{-1} & \text{if } |t| < \eta_\alpha \\ \bar{y} & \text{if } |t| \geq \eta_\alpha \end{cases} \qquad (2.16)$$

where η_α is the critical value of the Student's-t distribution $g(t)$ with $(n_x + n_y - 2)$ degrees of freedom (df) such that

$$\int_{-\eta_\alpha}^{\eta_\alpha} g(t)\, dt = 1 - \alpha. \qquad (2.17)$$

He obtained the cumulative distribution function and the frequency function of $\tilde{\hat{\mu}}_{y3}$ and then calculated $E(\tilde{\hat{\mu}}_{y3})$ and $MSE(\tilde{\hat{\mu}}_{y3})$.

Bennett also discussed testing the null hypothesis H_{2o}: $\mu_y = \mu_x$ versus the alternative hypothesis H_{2A}: $\mu_y \neq \mu_x$ after a preliminary F test of H_{1o}: $\sigma_y^2 = \sigma_x^2$ versus H_{1A}: $\sigma_y^2 \neq \sigma_x^2$. He used the same α level for both tests. He obtained

$$\Pr\left\{ \frac{|\bar{y} - \bar{x}|}{\tilde{s}\sqrt{n_x^{-1} + n_y^{-1}}} > \eta_\alpha \,\middle|\, H_{1o}: \sigma_y^2 = \sigma_x^2 \text{ is accepted} \right\}, \qquad (2.18)$$

i.e. the probability that the null hypothesis H_{2o}: $\mu_y = \mu_x$ is rejected given that H_{1o}: $\sigma_y^2 = \sigma_x^2$ is accepted.

In the fifth case he discussed a first preliminary test of H_{1o}: $\sigma_y^2 = \sigma_x^2$ versus H_{1A}: $\sigma_y^2 \neq \sigma_x^2$. If H_{1o} is accepted, a second preliminary test is made of H_{2o}: $\mu_y = \mu_x$ versus H_{2A}: $\mu_y \neq \mu_x$, using the test statistic t as in equation (2.15), where s^2 is the pooled estimator of σ^2 as in equation (2.14) and $\sigma_y^2 = \sigma_x^2 = \sigma^2$. The estimator of μ_y in this case is $\hat{\mu}_{y4}$,

where

$$\mu_{y4} = \begin{cases} \left[\dfrac{n_y \bar{y}}{s_y^2} + \dfrac{n_x \bar{x}}{s_x^2} \right] \left[\dfrac{n_y}{s_y^2} + \dfrac{n_x}{s_x^2} \right]^{-1} & \text{if } H_{1o} \text{ accepted and } |t| < \eta_\alpha \\ \bar{y} & \text{if } H_{1o} \text{ accepted and } |t| \geq \eta_\alpha \end{cases} . \tag{2.19}$$

s_x^2 and s_y^2 are the sample estimates of σ_x^2 and σ_y^2, respectively. He found $E(\tilde{\hat{\mu}}_{y4})$, given that H_{1o} is accepted, but discussed no estimation procedure if H_{1A} is accepted.

The formulas for bias and mean square error derived by Bennett in his 1952 paper are rather complicated, and it is difficult to tell by visual inspection how variation in the different components, such as α, σ^2, n_y, and n_x will effect the bias and mean square error. Bennett did not attempt to investigate this problem in his paper.

In a later paper Bennett (1956) considered two independent distributions $N(\mu_y, 1)$ and $N(\mu_x, 1)$ and independent sample means $\bar{y}$ and $\bar{x}$ based on n_y and n_x observations, respectively. It is desired to find a confidence interval for μ_y. The preliminary test statistic is

$$z = (\bar{y} - \bar{x})\ (1/n_y + 1/n_x)^{-\frac{1}{2}} \tag{2.20}$$

and the confidence interval is

$$(\hat{\mu}_{yL}, \hat{\mu}_{yU}) = \begin{cases} \left[\dfrac{n_y\bar{y} + n_x\bar{x}}{n_y + n_x} - \dfrac{\xi_\beta}{\sqrt{n_x + n_y}}, \dfrac{n_y\bar{y} + n_x\bar{x}}{n_y + n_x} + \dfrac{\xi_\beta}{\sqrt{n_x + n_y}} \right] & \text{if } |z| < \xi_\alpha \\ \left[\bar{y} - \dfrac{\xi_\beta}{\sqrt{n_y}}, \bar{y} + \dfrac{\xi_\beta}{\sqrt{n_y}} \right] & \text{if } |z| \geq \xi_\alpha \end{cases} \tag{2.21}$$

where

$$\int_{-\xi_\beta}^{\xi_\beta} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt = 1 - \beta. \tag{2.22}$$

If no preliminary test had been made and either confidence interval used all the time, the confidence coefficient would be $(100)(1 - \beta)\%$. Bennett then found the actual confidence coefficient of the interval, i.e. $\Pr[\tilde{\hat{\mu}}_{yL} \leq \mu_y \leq \tilde{\hat{\mu}}_{yU}]$. However, he gave no guidelines about the choice of α and β so as to achieve a preassigned confidence coefficient on the final confidence interval.

Kitagawa (1963) discussed several problems of estimation involving the use of preliminary tests of significance. He seems to be the only person to discuss this type of estimation procedure for dependent sample means. In investigating the use of interpenetrating subsamples in sample surveys in an earlier paper, Kitagawa (1954, p. 1125) said:

> "We have emphasized...that the true significance of inter-penetrating samples used in designs of sample surveys can

> be duly recognized only after we shall adopt some elaborate successive process of statistical inference. One of the ways which we may adopt under certain circumstances is to employ the pooling of data, that is, the estimation after the preliminary test of significance of the differences."

He was led into correlated sample means of the interpenetrating subsamples because in such sampling schemes the sample means often have common primary and secondary units. He also considered a normal model because "approximate normality can be established for estimates for subpopulation means, provided the sizes of subpopulations are sufficiently large." (Kitagawa, 1963, p. 152).

Kitagawa considered the bivariate normal distribution of $(\tilde{x}, \tilde{y})$ with parameters $E(\tilde{x}) = \mu_x$, $E(\tilde{y}) = \mu_y$, $V(\tilde{x}) = V(\tilde{y}) = \sigma^2$, and $\mathrm{cov}(\tilde{x}, \tilde{y}) = \rho\,\sigma^2$, where all the parameters are unknown. There is an a priori reason to prefer the y- observations to the x- observations if $\mu_y \neq \mu_x$. A bivariate random sample of size n is taken and the mean

$$\bar{d} = n^{-1} \sum_{i=1}^{n} d_i \tag{2.23}$$

and sample variance

$$s_d^2 = (n-1)^{-1} \sum_{\tilde{i}=1}^{n} (d_i - \bar{d})^2 \tag{2.24}$$

of the random variable d_i are computed, where

$$d_i = y_i - x_i \,. \tag{2.25}$$

The preliminary test statistic is defined as

$$t = \bar{d}\sqrt{n} \,/\, s_d \tag{2.26}$$

and the estimator is

$$\hat{\mu}_y = \begin{cases} (\bar{y} + \bar{x})/2 & \text{if } |t| < \eta_\alpha \\ \bar{y} & \text{if } |t| \geq \eta_\alpha \end{cases} \tag{2.27}$$

where η_α is the critical value of the Student-t distribution g(t) with (n - 1) df such that

$$\int_{-\eta_\alpha}^{\eta_\alpha} g(t)\, dt = 1 - \alpha. \tag{2.28}$$

Kitagawa found the cumulative distribution function of $\tilde{\hat{\mu}}_y$, as well as $E(\tilde{\hat{\mu}}_y)$ and $MSE(\tilde{\hat{\mu}}_y)$. He noted that the larger the correlation ρ between $\tilde{x}$ and $\tilde{y}$, the greater the discriminating power of the preliminary test. Ghosh (1949) and Mokashi (1949) had previously stated that the pooled estimator from interpenetrating subsamples has lower efficiency the larger the correlation. This is because they considered $V\,[(\tilde{\bar{x}} + \tilde{\bar{y}})/2] = \sigma^2 (1 + \rho^2)/2$. Kitagawa (1956) pointed out, however, that the use of interpenetrating subsamples should be judged not only on the efficiency of the joint estimator, but also by the discriminating power of the comparison between the samples. Indeed, the mean square error of the resulting estimator should take into account the fact that a preliminary test has been done. He noted that $MSE\ (\tilde{\hat{\mu}}_y)$ depends upon ρ through incomplete Beta functions and makes the evaluation of the effect of ρ on $MSE(\tilde{\hat{\mu}}_y)$ much more difficult than Ghosh and Mokashi claimed.

Kitagawa (1963) also discussed several other uses of preliminary tests and subsequent estimation, and his paper contains a good bibliography on the subject. Bancroft (1964) has also given several examples of the uses of preliminary tests, and his paper has a complete bibliography.

Huntsberger (1955) proposed a continuously weighted estimator which was suggested by the preliminary test and estimation scheme. He considered two independent, normally distributed sample means $\bar{x}$ and $\bar{y}$ with known variances $\sigma^2_{\bar{x}}$ and $\sigma^2_{\bar{y}}$. He defined

$$z = (\bar{y} - \bar{x})(\sigma^2_{\bar{y}} + \sigma^2_{\bar{x}})^{-\frac{1}{2}} \tag{2.29}$$

and then proposed the estimator

$$\hat{\mu}_{yH} = [\phi(z)]\bar{y} + [1 - \phi(z)]\,(\sigma^2_{\bar{x}}\ \bar{y} + \sigma^2_{\bar{y}}\ \bar{x})(\sigma^2_{\bar{x}} + \sigma^2_{\bar{y}})^{-1} \tag{2.30}$$

where $\phi(z)$ is any function of z satisfying some mild restrictions. Note that $\hat{\mu}_{yH}$ is a weighted sum of the unpooled estimator $\bar{y}$ and the pooled estimator $(\sigma^2_{\bar{x}}\ \bar{y} + \sigma^2_{\bar{y}}\ \bar{x})(\sigma^2_{\bar{x}} + \sigma^2_{\bar{y}})^{-1}$, where the weights sum to one. Huntsberger defined δ as

$$\delta = (\mu_y - \mu_x)(\sigma^2_{\bar{y}} + \sigma^2_{\bar{x}})^{-\frac{1}{2}}, \tag{2.31}$$

replaced the weight function $\phi(z)$ by A, where A is a function of δ, and found that $MSE(\tilde{\hat{\mu}}_{yH}) = E(\tilde{\hat{\mu}}_{yH} - \mu_y)^2$ is minimized for $A = \delta^2\ (1 + \delta^2)^{-1}$. Since $E(\tilde{z}) = \delta$, he proposed letting

$$\phi(z) = z^2\ (1 + z^2)^{-1}. \tag{2.32}$$

For this choice of $\phi(z)$, note that as $z \longrightarrow 0$,

$$\hat{\mu}_{yH} \longrightarrow (\sigma^2_{\bar{x}}\ \bar{y} + \sigma^2_{\bar{y}}\ \bar{x})(\sigma^2_{\bar{x}} + \sigma^2_{\bar{y}})^{-1}$$

and that as $z \longrightarrow \infty$, $\hat{\mu}_{yH} \longrightarrow \bar{y}$. He derived $MSE(\tilde{\hat{\mu}}_{yH})$ with $\phi(z)$ as in equation (2.32) and found that $MSE(\tilde{\hat{\mu}}_{yH})$, as a function of δ, has a larger minimum and a smaller maximum than the corresponding preliminary test procedure. The preliminary test procedure yields either the pooled or unpooled estimator whereas Huntsberger's procedure always yields a weighted average of the two estimators, and thus it is a more conservative estimation procedure.

Some other authors have discussed the pooling of information to estimate a mean μ, but have assumed a priori that all the parameter means are equal. For example, Halperin (1961) discussed the estimation of the parameter μ, where the vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p)'$ has a multivariate normal distribution with unknown mean vector $(\mu, \mu, \ldots, \mu)'$ and unknown variance-covariance matrix Σ. He also discussed the case where $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p$ are independent, and independent samples with unequal sample sizes and unknown, unequal variances are to be pooled to obtain an estimate of the common mean μ. Zacks (1966) discussed the estimation of the common mean of two independent normal distributions when the variances are unknown and the sample sizes are equal. An estimator efficient in large samples is

$$\hat{\mu} = \left[\frac{\bar{y}}{s_y^2} + \frac{\bar{x}}{s_x^2} \right] \left[\frac{1}{s_y^2} + \frac{1}{s_x^2} \right]^{-1} \tag{2.33}$$

where s_x^2 and s_y^2 are the sample variances. He also suggested doing a preliminary test of H_o: $\sigma_y^2 = \sigma_x^2$ versus H_1: $\sigma_y^2 > \sigma_x^2$ and H_{-1}: $\sigma_y^2 < \sigma_x^2$. His estimator is $(\bar{y} + \bar{x})/2$ if H_o is accepted, $\bar{x}$ if H_1 is accepted, and $\bar{y}$ if H_{-1} is accepted.

Many other papers discuss classical estimation of a common mean, but since this investigation is not willing to assume a priori that all the means are equal, no more papers of this type shall be discussed.

Mosteller (1948) seems to be the only person to discuss the Bayesian approach to the preliminary test and estimation problem. Other authors have discussed pooling data from a Bayesian point of view, but they have assumed a priori that all the different means are equal. For example, Jeffreys (1961) discussed the combination of independent, normally distributed estimates $\tilde{\bar{y}}_i$ with sample variances $\tilde{s}_i^2$ based on υ_i df assuming that $E(\tilde{\bar{y}}_i) = \mu$ for all i. He took the prior distributions on $\tilde{\mu}$ and log $\tilde{\sigma}_i$ to be diffuse and independent of each other. The posterior distribution of $\tilde{\mu}$, given the sample data, is a product of independent t-distributions, which is not a t-distribution itself. Jeffreys suggested reducing the posterior distribution to an approximate t-distribution and then using the posterior mean as the estimator of $\tilde{\mu}$.

Geisser (1965a) discussed the case of combining correlated estimates via the Bayesian approach when the vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p)'$ has the multivariate normal distribution with unknown mean vector $(\mu, \mu, \ldots, \mu)'$ and unknown variance-covariance matrix Σ. He also considered non-informative prior distributions on $\tilde{\mu}$ and $\tilde{\Sigma}$ and took as the estimator of μ the mean of the posterior distribution of $\tilde{\mu}$, given the sample data, i.e.

$$\hat{\mu} = \frac{e' \, S^{-1} \, \bar{x}}{e' \, S^{-1} \, e} \qquad (2.34)$$

where $e' = (1, 1, \ldots, 1)'$, $\bar{x}' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$, and S is the sample variance-covariance matrix. Geisser also discussed other priors on $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$, the most important one being the natural conjugate Bayes density which puts a mormal prior on $\underset{\sim}{\mu}$ and a Wishart prior on $\underset{\sim}{\Sigma}$, the two priors being independent.

Again, there may be other papers which discuss the Bayesian estimation of a common mean by assuming a priori that all the means are equal, but that is not the primary interest here. These cases are mentioned merely to illustrate another approach to the problem of pooling means.

III. A GENERAL PRELIMINARY TEST PROCEDURE FOR ESTIMATING THE MEAN OF A NORMAL POPULATION

A. Statement of Problem

In this chapter a scheme of estimating one of the means of a bivariate normal population is discussed when observations on both random variables $\tilde{x}$ and $\tilde{y}$ are available. A very general, two stage procedure which involves a preliminary test of significance is proposed, and then the bias and mean square error of the procedure are derived. In succeeding chapters various special cases of the procedure are considered in more detail.

B. The Sampling Scheme and Its Motivation

A simple random bivariate sample of size n is selected from the infinite population of random variables $(\tilde{x}, \tilde{y})$ which follow a bivariate probability distribution. In addition an independent simple random sample of n_x additional observations is taken on $\tilde{x}$, ignoring the corresponding measurement on $\tilde{y}$ for these variables. Likewise, n_y additional independent observations on $\tilde{y}$ are taken, ignoring the corresponding measurement on $\tilde{x}$. In the sample now are $(n + n_x)$ observations on $\tilde{x}$ and $(n + n_y)$ observations on $\tilde{y}$. The only existing correlation between observations is between the elements of the pair (x_i, y_i), i=1, 2,..., n, where (x_i, y_i) belongs to the bivariate sample of size n.

There are several reasons for considering this type of sample. In many cases n_x and n_y will be zero, resulting in a regular bivariate sample. However, some extra observations may be available on one or both of the random variables. For example, data on some attribute from a census could correspond to a random sample of size $(n + n_x)$ on the random variable $\tilde{x}$, whereas data from a post-enumeration survey on the same attribute could correspond to a random sample of size n on the random variable $\tilde{y}$. In most cases every element in the sample on $\tilde{y}$ would be matched to an element in the sample on $\tilde{x}$. In this case, $n_y = 0$. The two samples would be pooled if the experimenter felt that they were really measuring the same attribute, and he would probably assume that the post-enumeration survey provided the most accurate estimate. Also, the experimenter may be presented with a size n bivariate sample, and he may wish to do some additional independent sampling on one or both of the variables before he makes his preliminary test of significance. Finally, the result of an optimal allocation may suggest a sample of this type.

At this stage a preliminary test of the null hypothesis $H_o: \mu_y = \mu_x$ versus the alternative hypothesis $H_A: \mu_y \neq \mu_x$ is made. If $H_o: \mu_y = \mu_x$ is accepted, a further simple random sample of m_x observations is taken on $\tilde{x}$ only, and the estimator μ_y is some appropriately weighted average of all the observations on $\tilde{x}$ and $\tilde{y}$. If, however, $H_A: \mu_y \neq \mu_x$ is accepted, a further simple random sample of m_y observations is taken on $\tilde{y}$ only, and the estimator μ_y is the mean of all the observations on $\tilde{y}$.

The primary motivation for the second stage sampling scheme is the cost of collecting the data. It is assumed that the variable of primary interest is $\tilde{y}$, but that $\tilde{y}$ is more expensive to measure than $\tilde{x}$. Thus, if $\mu_y = \mu_x$, observations on $\tilde{x}$ furnish an estimate of μ_y at a cheaper cost. This is why, if H_o: $\mu_y = \mu_x$ is accepted, $\tilde{x}$'s are sampled in the second stage of sampling and a pooled estimator of μ_y is used. If the alternative hypothesis H_A: $\mu_y \neq \mu_x$ is accepted, it is concluded that μ_y and μ_x differ significantly, and thus the observations on $\tilde{x}$ are not used in a pooled estimator. This is why, when H_A: $\mu_y \neq \mu_x$ is accepted, additional $\tilde{y}$'s are sampled at the second stage of sampling and the final estimator is simply the mean of the observations on $\tilde{y}$. A secondary motivation is that the experimenter may strongly prefer observations on $\tilde{y}$ instead of $\tilde{x}$ for other than financial reasons. Thus again he would sample additional $\tilde{x}$'s only if there is strong evidence that the two random variables measure the same thing.

As a particular application of the two stage feature, the experimenter may be presented with data on $\tilde{x}$ and $\tilde{y}$, but he is unsure whether $\mu_y = \mu_x$ and, in addition, he knows that he does not have enough data to obtain the precision he wants for the estimation of μ_y. In this case, a preliminary test of significance followed by the second stage sampling would be appropriate.

The sampling and estimation scheme is represented in Figure 3.1. It

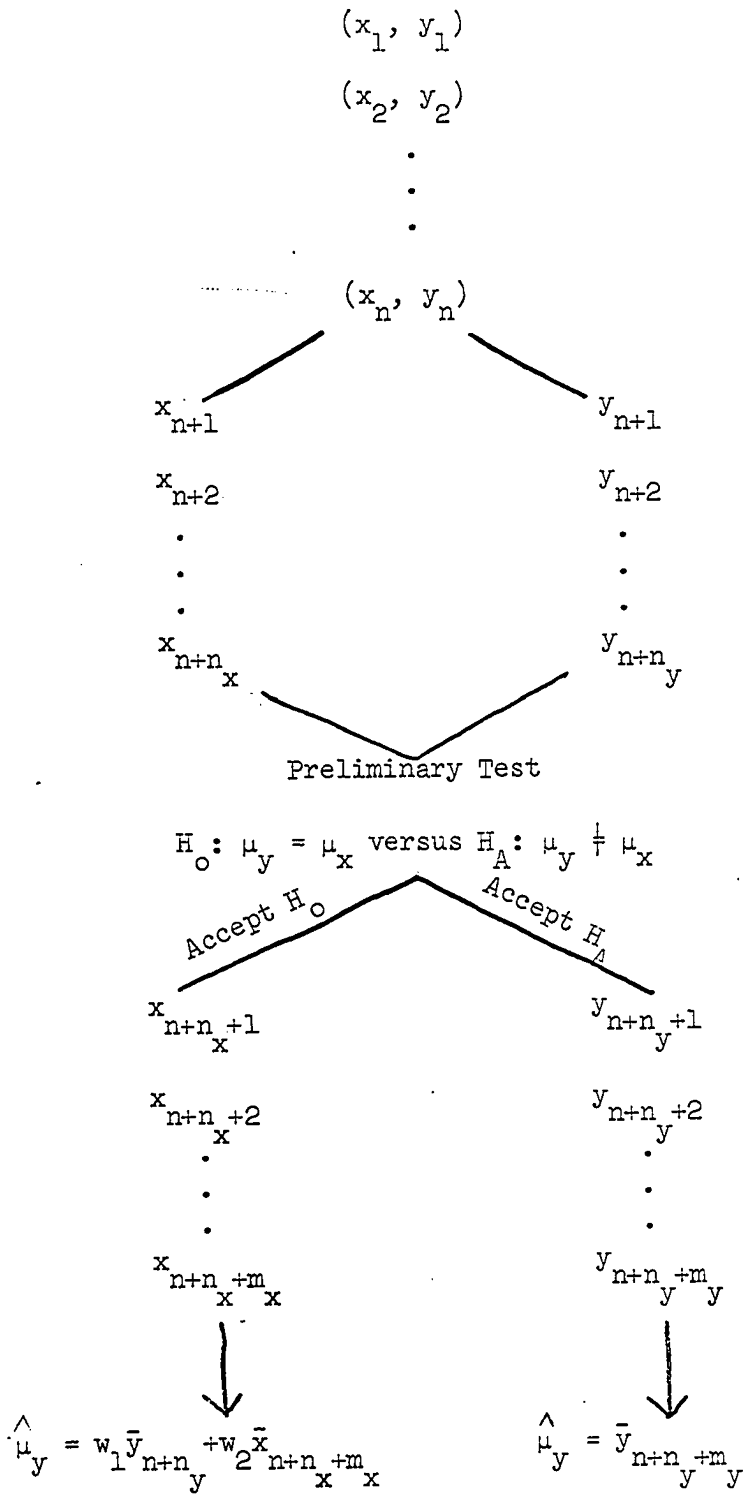

Figure 3.1. The two stage sampling scheme

could be argued that the use of only the observations on $\tilde{y}$ when $H_0: \mu_y = \mu_x$ is rejected is not very efficient since the observations on $\tilde{x}$ are not used in the resultant estimator. If $H_A: \mu_y \neq \mu_x$ is accepted, then a regression estimator would use all of the available data. This approach is discussed in Chapter VI. However, there are several situations where a regression estimator could not be used very advantageously, and the scheme proposed in this chapter would be the logical alternative. The most obvious case where a regression estimator is useless is the special case where $\tilde{x}$ and $\tilde{y}$ are independent. Second, if only a bivariate sample of size n is available, i.e. $n_x = n_y = m_x = m_y = 0$, then a regression estimator cannot be used since no observations are available on $\tilde{x}$ other than those matched with observations on $\tilde{y}$. Third, if the number of additional independent observations on $\tilde{x}$ beyond those in the bivariate sample is small, then the regression term will probably be near zero. Fourth, in cases where the covariance matrix is unknown, the regression coefficient will have to be estimated from the bivariate sample of size n. Thus, if n is very small, the regression estimator probably should not be used because of the unreliable estimation of the regression coefficient. For these reasons, the sampling scheme described in this chapter has relevant applications.

C. Mathematical Specification of the Preliminary Test Estimation Procedure

The bivariate random variable $(\tilde{x}, \tilde{y})$ is assumed to follow the bivariate normal distribution with mean

$$E\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \tag{3.3.1}$$

and covariance matrix

$$V\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \tag{3.3.2}$$

The bivariate normal distribution is assumed for simplicity, but it would be useful to investigate the behavior of this preliminary test scheme when the data are not selected from a bivariate normal distribution, but are, in fact, from some other distribution, such as bivariate gamma. The bias and mean square error as reported in this chapter would not be correct for non-normal data, but they may provide reasonable approximations to the true bias and mean square error if the procedure is robust. An alternative to investigating robustness in this manner would be to derive the bias and mean square error for a preliminary test procedure with the assumption that the data come from, say, a bivariate gamma distribution.

To begin the investigation of this procedure, the elements of the covariance matrix (i.e. ρ, σ_x^2, σ_y^2) are assumed to be known. If this assumption is removed, the derivation of bias and mean square error for this two stage procedure is further complicated. For example, the preliminary test involves the comparison of the estimators of μ_y and μ_x. If the covariance matrix is known, then this comparison follows the $N(0,1)$ distribution. However, if the covariance matrix is unknown, several problems arise. First, consider the special case where $\tilde{x}$ and $\tilde{y}$ are independent, i.e. when $\rho = 0$. Bennett (1952) has considered the further special case where $\sigma_y^2 = \sigma_x^2 = \sigma^2$ with σ^2 unknown, which then admits a Student-t test as the preliminary test statistic. However, if $\sigma_y^2 \neq \sigma_x^2$ are both unknown, then the test of H_o: $\mu_y = \mu_x$ versus H_A: $\mu_y \neq \mu_x$ involves distributional problems such as the Behrens-Fisher problem, etc. If, further, ρ is unknown, then additional complications arise. Kitagawa (1963) considered the special case where ρ and $\sigma_y^2 = \sigma_x^2 = \sigma^2$ are unknown and then analyzed paired differences on a bivariate random sample of size n. For these reasons, the properties of the procedure are derived with known covariance matrix, with the hope that these properties will not change drastically when the covariance matrix is unknown. A preliminary investigation of the procedure with unknown covariance matrix is reported in Chapter VIII.

The covariance structure of the observations in Figure 3.1 can be expressed mathematically as:

$$\begin{aligned}
\mathrm{cov}(\tilde{x}_i, \tilde{y}_j) &= \rho\sigma_x\sigma_y && \text{for } i=j,\ i=1, 2, \ldots, n \\
\mathrm{cov}(\tilde{x}_i, \tilde{y}_j) &= 0 && \text{otherwise} \\
\mathrm{cov}(\tilde{x}_i, \tilde{x}_j) &= 0 && \text{for } i \neq j \\
\mathrm{cov}(\tilde{y}_i, \tilde{y}_j) &= 0 && \text{for } i \neq j
\end{aligned} \tag{3.3.3}$$

The various means of interest are defined below.

$$\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$$

$$\bar{y}_n = n^{-1} \sum_{i=1}^{n} y_i$$

$$\bar{y}_{n_y} = n_y^{-1} \sum_{i=n+1}^{n+n_y} y_i$$

$$\bar{x}_{n_x} = n_x^{-1} \sum_{i=n+1}^{n+n_x} x_i$$

$$\bar{y}_{m_y} = m_y^{-1} \sum_{i=n+n_y+1}^{n+n_y+m_y} y_i \qquad (3.3.4)$$

$$\bar{x}_{m_x} = m_x^{-1} \sum_{i=n+n_x+1}^{n+n_x+m_x} x_i$$

$$\bar{x}_{n+n_x} = (n+n_x)^{-1} (n \cdot \bar{x}_n + n_x \bar{x}_{n_x})$$

$$\bar{y}_{n+n_y} = (n+n_y)^{-1} (n \bar{y}_n + n_y \bar{y}_{n_y})$$

$$\bar{x}_{n+n_x+m_x} = (n+n_x+m_x)^{-1} (n \bar{x}_n + n_x \bar{x}_{n_x} + m_x \bar{x}_{m_x})$$

$$\bar{y}_{n+n_y+m_y} = (n+n_y+m_y)^{-1} (n \bar{y}_n + n_y \bar{y}_{n_y} + m_y \bar{y}_{m_y})$$

If H_o: $\mu_y = \mu_x$ is accepted, the estimator of μ_y is $\hat{\mu}_y = v$, where

$$v = w_1 \bar{y}_{n+n_y} + w_2 \bar{x}_{n+n_x+m_x} \qquad (3.3.5)$$

and w_1 and w_2 are two constants such that $w_1 + w_2 = 1$. If H_A: $\mu_y \neq \mu_x$ is accepted, the estimator of μ_y is $\hat{\mu}_y = u$, where

$$u = \bar{y}_{n+n_y+m_y} . \tag{3.3.6}$$

Let the cost of measuring $\tilde{x}$ for one unit be C_x, the cost of measuring $\tilde{y}$ for one unit be C_y, and the cost of measuring a correlated pair $(\tilde{x}, \tilde{y})$ be C_{xy}. It is assumed that

$$C_x < C_y < C_{xy} < C_x + C_y . \tag{3.3.7}$$

Although $C_x < C_y$, it will usually be the case that $\sigma_x^2 > \sigma_y^2$, i.e. the cheaper observations are more variable. So, to obtain minimum mean square error for a fixed budget, it is necessary to realize that for a given cost more observations on $\tilde{x}$ can be measured than on $\tilde{y}$, but their variance may possibly exceed the variance of a smaller number of observations on $\tilde{y}$ obtained at the same cost. In an extreme case, the best strategy might be to sample no $\tilde{x}$'s at all, even though it is known that $\mu_y = \mu_x$.

D. Choice of Test Statistic for the Preliminary Test

To test the null hypothesis H_o: $\mu_y = \mu_x$ against the alternative hypothesis H_A: $\mu_y \neq \mu_x$, the statistic z is used, where

$$z = \bar{y}_{n+n_y} - \bar{x}_{n+n_x} . \tag{3.4.1}$$

The variance of $\tilde{z}$ is

$$V(\tilde{z}) = \sigma_z^2 = \frac{\sigma_y^2}{n+n_y} + \frac{\sigma_x^2}{n+n_x} - 2 \operatorname{cov}(\bar{y}_{n+n_y}, \bar{x}_{n+n_x}) . \tag{3.4.2}$$

Now,

$$\operatorname{cov}(\tilde{\bar{y}}_{n+n_y}, \tilde{\bar{x}}_{n+n_x}) = \operatorname{cov}\left[\sum_{i=1}^{n+n_y} \frac{\tilde{y}_i}{n+n_y}, \sum_{j=1}^{n+n_x} \frac{\tilde{x}_j}{n+n_x}\right]$$

$$= \frac{1}{(n+n_x)(n+n_y)} \sum_{i=1}^{n+n_y} \operatorname{cov}\left[\tilde{y}_i, \sum_{j=1}^{n+n_x} \tilde{x}_j\right]. \tag{3.4.3}$$

By the covariance specifications in equation (3.3.3),

$$\begin{aligned} \operatorname{cov}\left[\tilde{y}_i, \sum_{j=1}^{n+n_x} \tilde{x}_j\right] &= \rho\sigma_x\sigma_y \quad \text{for } i=1, 2, \ldots, n \\ \operatorname{cov}\left[\tilde{y}_i, \sum_{j=1}^{n+n_x} \tilde{x}_j\right] &= 0 \qquad \text{for } i=n+1, \ldots, n+n_y. \end{aligned} \tag{3.4.4}$$

Substituting equation (3.4.4) into (3.4.3) yields

$$\operatorname{cov}(\tilde{\bar{y}}_{n+n_y}, \tilde{\bar{x}}_{n+n_x}) = (n+n_x)^{-1} (n+n_y)^{-1} n\, \rho\sigma_x\sigma_y \tag{3.4.5}$$

and thus

$$V(\tilde{z}) = \frac{\sigma_y^2}{n+n_y} + \frac{\sigma_x^2}{n+n_x} - \frac{2n\rho\sigma_x\sigma_y}{(n+n_x)(n+n_y)}\ . \tag{3.4.6}$$

If the correlation ρ is near one, then better precision on the preliminary test might be obtained by using the test statistic z' instead of z, where

$$z' = \bar{y}_n - \bar{x}_n \tag{3.4.7}$$

and

$$V(\tilde{z}') = n^{-1} (\sigma_y^2 + \sigma_x^2 - 2\rho\sigma_x\sigma_y). \tag{3.4.8}$$

This is because

$V(\tilde{z}) < V(\tilde{z}')$ if, and only if,

$$\rho < \frac{\sigma_y^2 n_y(n+n_x) + \sigma_x^2 n_x(n+n_y)}{2\sigma_x\sigma_y\,(nn_x + nn_y + n_x n_y)} . \tag{3.4.9}$$

Obviously, inequality (3.4.9) is satisfied if $\rho \leq 0$. However, if, for example, $n_y = n$, $n_x = 2n$, $\sigma_y^2 = 4$, and $\sigma_x^2 = 9$, then inequality (3.4.9) becomes $\rho < 4/5$. Thus, if $\rho = 5/6$, then $V(z) > V(z')$. In general, inequality (3.4.9) is satisfied unless ρ is close to one.

It appears that the use of z' instead of z modifies the bias and mean square error only slightly unless n_x and/or n_y are very large with respect to n. Since z has the best precision for most possible values of ρ, and since using all of the available information is a good strategy most of the time, the statistic z will be used to test H_o: $\mu_y = \mu_x$ against H_A: $\mu_y \neq \mu_x$. For most applications the correlation will be zero or moderately positive, and thus inequality (3.4.9) will be satisfied most of the time.

E. A Lemma on the Distribution of Linear Combinations of Normally Distributed Random Variables

In deriving the bias and mean square error of the two stage preliminary test procedure, it is necessary to know the joint distribution of random variables such as $\tilde{\bar{y}}_{n+n_y}$ and $\tilde{\bar{x}}_{n+n_x}$. An intelligent guess would be that the joint distribution could hardly be anything other than bivariate normal since the original random variables are selected from a bivariate normal distribution. However, the sampling scheme involved in this procedure is not bivariate sampling in the usual sense, and bivariate normality does not follow directly from sampling theorems on the multivariate normal distribution. However, bivariate normality can be proved by standard methods, and Lemma 3.1 below is stated without proof.

Lemma 3.1. Let the random variables $\tilde{x}_i$, i=1, 2,..., $n+n_x$ and $\tilde{y}_j$, j=1, 2,..., $n+n_y$, be selected from a bivariate normal distribution by the sampling method described in sections B and C of this chapter. Then the joint distribution of any new random variables which are linear combinations of the random variables $\tilde{x}_i$ and $\tilde{y}_j$ is the multivariate normal distribution.

F. A Lemma on Conditional Expectation

In the derivation of bias and mean square error of the two stage preliminary test procedure, it is necessary to evaluate conditional expectations of the type $E\left[\tilde{x}\tilde{y} \,\middle|\, |\tilde{x}| < c\right]$, where $\tilde{x}$ and $\tilde{y}$ are correlated and c is a real, positive number. Lemma 3.2 and its corollary present formulas for some conditional expectations which will be used repeatedly throughout this dissertation. They can be proved by standard methods and are thus stated without proof.

Lemma 3.2. Let $(\tilde{x},\tilde{y})$ have a bivariate distribution with joint density function $f(x,y)$, marginal density function $h(y)$ for $\tilde{y}$, and conditional density function $g(x|y)$ for $\tilde{x}|y$. Let $c > 0$ be a real number. Let $b(\tilde{x})$ and $d(\tilde{y})$ be random variables which are continuous functions of the random variables $\tilde{x}$ and $\tilde{y}$, respectively. Then

$$E\left[b(\tilde{x})\ d(\tilde{y}) \;\middle|\; |\tilde{y}| < c \right] = \frac{\int_{-c}^{c} d(y)\ E[b(\tilde{x}) \mid y]\ h(y)\ dy}{\Pr\left[\,|\tilde{y}| < c\right]} \tag{3.6.1}$$

Corollary to Lemma 3.2. If the hypothesis to Lemma 3.2 holds, then

$$E\left[b(\tilde{x})\ d(\tilde{y}) \;\middle|\; |\tilde{y}| \geq c \right] = \frac{E[b(\tilde{x})d(\tilde{y})] - \int_{-c}^{c} d(y)\ E[b(\tilde{x}) \mid y]\ h(y)\ dy}{\Pr\left[\,|\tilde{y}| \geq c\right]}. \tag{3.6.2}$$

G. Bias of the Two Stage, Preliminary Test Estimation Procedure

$E(\tilde{\hat{\mu}}_y)$ is now derived, where $\hat{\mu}_y$ is the estimator described in section B of this chapter and represented in Figure 3.1. In section D it was decided to use $z = \bar{y}_{n+n_y} - \bar{x}_{n+n_x}$ as the test statistic to test the null hypothesis H_o: $\mu_y = \mu_x$ versus the alternative hypothesis H_A: $\mu_y \neq \mu_x$. By Lemma 3.1 of this chapter, $\tilde{z}$ is normally distributed with mean

$$E(\tilde{z}) = \mu_y - \mu_x = \Delta \tag{3.7.1}$$

and variance $V(\tilde{z}) = \sigma_z^2$ given in equation (3.4.6). The density function of $\tilde{z}$ is $h(z)$, where

$$h(z) = (2\pi)^{-\frac{1}{2}}\ \sigma_z^{-1} e^{\frac{-(z-\Delta)^2}{2\sigma_z^2}}\ . \tag{3.7.2}$$

To make a preliminary test of size α, a critical value ξ_α is chosen from the N(0, 1) table such that

$$\int_{-\xi_\alpha}^{\xi_\alpha} \phi\,(t)\ dt = 1 - \alpha, \tag{3.7.3}$$

where

$$\phi(t) = (2\pi)^{-\frac{1}{2}} e^{-t^2/2} \tag{3.7.4}$$

Thus, if $|z| < \xi_\alpha \sigma_z$, H_o is accepted; if $|z| \geq \xi_\alpha \sigma_z$, H_A is accepted.

Proceeding with the expectation of $\hat{\mu}_y$,

$$\begin{aligned} E(\tilde{\hat{\mu}}_y) &= E[\tilde{\hat{\mu}}_y \mid \text{Accept } H_o] \Pr[\text{Accept } H_o] \\ &\quad + E[\tilde{\hat{\mu}}_y \mid \text{Accept } H_A] \Pr[\text{Accept } H_A] \\ &= E[\tilde{v} \mid |\tilde{z}| < \xi_\alpha \sigma_z] \Pr[|\tilde{z}| < \xi_\alpha \sigma_z] \\ &\quad + E[\tilde{u} \mid |\tilde{z}| \geq \xi_\alpha \sigma_z] \Pr[|\tilde{z}| \geq \xi_\alpha \sigma_z] \end{aligned} \tag{3.7.5}$$

where u and v are the unpooled and pooled estimators defined in equations (3.3.6) and (3.3.5).

By Lemma 3.1 each of the joint distributions of $(\tilde{\bar{y}}_{n+n_y}, \tilde{z})$ and $(\tilde{\bar{x}}_{n+n_x}, \tilde{z})$ follows the bivariate normal distribution. By a similar argument to that used in Lemma 3.1, $(\tilde{\bar{y}}_{n+n_y+m_y}, \tilde{z})$ and $(\tilde{\bar{x}}_{n+n_x+m_x}, \tilde{z})$ each follows the bivariate normal distribution. Then, by Lemma 3.2 of this chapter and its corollary,

$$E(\tilde{\hat{\mu}}_y) = E(\tilde{\bar{y}}_{n+n_y+m_y}) + \int_{|\xi|<\xi_\alpha\sigma_z} E(\tilde{G} \mid z)\, h(z)\, dz \tag{3.7.6}$$

where h(z) is defined in equation (3.7.2) and

$$G = w_1 \bar{y}_{n+n_y} + w_2 \bar{x}_{n+n_x+m_x} - \bar{y}_{n+n_y+m_y}. \tag{3.7.7}$$

Noting that $\bar{x}_{n+n_x} = (\bar{y}_{n+n_y} - z)$ and using the substitutions from equation (3.3.4),

$$G = \left[w_1 + \frac{w_2(n+n_x)}{N_x} - \frac{(n+n_y)}{N_y}\right]\bar{y}_{n+n_y} - \frac{w_2(n+n_x)z}{N_x}$$

$$+ \frac{w_2 m_x \bar{x}_{m_x}}{N_x} - \frac{m_y}{N_y}\bar{y}_{m_y} \qquad (3.7.8)$$

where

$$N_x = n + n_x + m_x$$

$$N_y = n + n_y + m_y. \qquad (3.7.9)$$

Given that $|\tilde{z}| < \xi_\alpha \sigma_z$, then $\tilde{\tilde{x}}_{m_x}$ and $\tilde{z}$ are independent. Likewise, given that $|\tilde{z}| \geq \xi_\alpha \sigma_z$, then $\tilde{\tilde{y}}_{m_y}$ and $\tilde{z}$ are independent. Thus,

$$E(\tilde{\tilde{x}}_{m_x} \mid z) = E(\tilde{\tilde{x}}_{m_x}) = \mu_x = \mu_y - \Delta \qquad (3.7.10)$$

and

$$E(\tilde{\tilde{y}}_{m_y} \mid z) = E(\tilde{\tilde{y}}_{m_y}) = \mu_y \qquad (3.7.11)$$

By Lemma 3.1 and standard multivariate normal theory, the joint distribution of $\tilde{\tilde{y}}_{n+n_y}$ and $\tilde{z}$ is bivariate normal with mean

$$E\begin{bmatrix}\tilde{\tilde{y}}_{n+n_y} \\ \tilde{z}\end{bmatrix} = \begin{bmatrix}\mu_y \\ \Delta\end{bmatrix} \qquad (3.7.12)$$

and covariance matrix

$$V\begin{bmatrix}\tilde{\tilde{y}}_{n+n_y} \\ \tilde{z}\end{bmatrix} = \begin{bmatrix}\dfrac{\sigma_y^2}{n+n_y} & k_1 \\ k_1 & \sigma_z^2\end{bmatrix} \qquad (3.7.13)$$

where

$$k_1 = \sigma_y^2/(n+n_y) - n\rho\sigma_x\sigma_y/\left\{(n+n_x)(n+n_y)\right\}$$

$$k_2 = \sigma_x^2/(n+n_x) - n\rho\sigma_x\sigma_y/\left\{(n+n_x)(n+n_y)\right\} \qquad (3.7.14)$$

$$k_1 + k_2 = \sigma_z^2$$

By standard normal theory, the conditional distribution of $\tilde{\tilde{y}}_{n+n_y}$, given z, is normal with mean

$$E(\tilde{\tilde{y}}_{n+n_y} \mid z) = \mu_y + k_1 (z-\Delta)/(k_1+k_2) \tag{3.7.15}$$

and variance

$$V(\tilde{\tilde{y}}_{n+n_y} \mid z) = \frac{\sigma_y^2}{n+n_y} - \frac{k_1^2}{k_1+k_2}\,. \tag{3.7.16}$$

Also,

$$V(\tilde{\tilde{y}}_{n+n_y} \mid z) = \frac{\sigma_x^2\,\sigma_y^2\left[1 - \dfrac{\rho^2\, n^2}{(n+n_x)(n+n_y)}\right]}{(k_1+k_2)(n+n_x)(n+n_y)}\quad. \tag{3.7.17}$$

Thus, taking G as in equation (3.7.8) and evaluating $E(\tilde{G}|z)$ by using equations (3.7.10), (3.7.11), and (3.7.15), it follows that

$$E(\tilde{G}|z) = \left[w_1 + \frac{w_2(n+n_x)}{N_x} - \frac{(n+n_y)}{N_y}\right]\left[\mu_y + \frac{k_1(z-\Delta)}{k_1+k_2}\right]$$
$$- \frac{w_2(n+n_x)z}{N_x} + \frac{w_2 m_x(\mu_y-\Delta)}{N_x} - \frac{m_y}{N_y}\,\mu_y$$

i.e.

$$E(\tilde{G}|z) = \left[\frac{m_y}{N_y} - \frac{w_2 m_x}{N_x}\right]\frac{k_1(z-\Delta)}{(k_1+k_2)} - \frac{w_2(n+n_x)z}{N_x} - \frac{w_2 m_x \Delta}{N_x}\,. \tag{3.7.18}$$

Substituting equation (3.7.18) into (3.7.6), noting that $E(\tilde{\tilde{y}}_{n+n_y+m_y}) = \mu_y$, and using the change of variable $t = (z-\Delta)/\sigma_z$, equation (3.7.6) yields

$$E(\tilde{\hat{\mu}}_y) = \mu_y + \int_{-\xi\,-\delta}^{\xi\,-\delta} \phi(t)\left[\left\{\frac{m_y}{N_y} - \frac{w_2 m_x}{N_x}\right\}\frac{k_1 t}{\sigma_z} - \frac{w_2(n+n_x)(t\sigma_z+\Delta)}{N_x}\right.$$
$$\left. - \frac{w_2 m_x \Delta}{N_x}\right] dt \tag{3.7.19}$$

where $\phi(t)$ is defined in equation (3.7.4) and

$$\delta = \Delta/\sigma_z . \tag{3.7.20}$$

Combining terms in equation (3.7.19) yields

$$E(\tilde{\hat{\mu}}_y) = \mu_y + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \phi(t) \left[-w_2\Delta + \frac{t}{\sigma_z} \left\{ -w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 k_2 m_x}{N_x} \right\} \right] dt. \tag{3.7.21}$$

From the definition of bias of an estimator,

$$\text{Bias } (\tilde{\hat{\mu}}_y) = E(\tilde{\hat{\mu}}_y) - \mu_y, \tag{3.7.22}$$

and considering Bias $(\tilde{\hat{\mu}}_y)$ as a function of δ, i.e. $B(\delta)$, equation (3.7.21) yields

$$B(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \phi(t) \left[-w_2\delta\sigma_z + \frac{t}{\sigma_z} \left\{ -w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 k_2 m_x}{N_x} \right\} \right] dt. \tag{3.7.23}$$

A few properties of $B(\delta)$ are now investigated. First,

$$B(-\delta) = \int_{-\xi_\alpha+\delta}^{\xi_\alpha+\delta} \phi(t) \left[w_2\delta\sigma_z + \frac{t}{\sigma_z} \left\{ -w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 k_2 m_x}{N_x} \right\} \right] dt. \tag{3.7.24}$$

Making the change of variable $t = -y$ and noting that $\phi(-y) = \phi(y)$,

$$B(-\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha+\delta} \phi(y) \left[w_2\delta\sigma_z + \frac{y}{\sigma_z} \left\{ -w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 k_2 m_x}{N_x} \right\} \right] dy$$

i.e.

$$B(-\delta) = -B(\delta). \tag{3.7.25}$$

$B(\delta)$ is thus an odd function of δ, and it is necessary to investigate $B(\delta)$ only for $\delta \geq 0$.

In order to evaluate $B(\delta)$ at $\delta = 0$, note that

$$\int_{-\xi_\alpha}^{\xi_\alpha} t\, \phi(t)\, dt = 0 \qquad (3.7.26)$$

and then, from equation (3.7.23), it follows easily that

$$B(0) = 0. \qquad (3.7.27)$$

Thus, the estimator $\hat{\mu}_y$ is unbiased if the population means μ_y and μ_x are equal.

In order to evaluate $B(\delta)$ as $\delta \longrightarrow \infty$, it is necessary to first investigate some lemmas on the limit of certain functions.

Lemma 3.3. Let n be a non-negative integer, ξ_α be a real, finite number, and

$$\phi(t) = (2\pi)^{-\frac{1}{2}}\, e^{-\frac{t^2}{2}}.$$

Then

$$\lim_{\delta \longrightarrow \infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n\, \phi(t)\, dt = 0.$$

Proof: Consider first n to be odd. Then, by repeated integration by parts,

$$\int t^n\, \phi(t)\, dt = -t^{n-1}\phi(t) - (n-1)t^{n-3}\phi(t) - (n-1)(n-3)t^{n-5}\phi(t)$$

$$- \ldots - (n-1)(n-3) \ldots 2\phi(t) \qquad (3.7.28)$$

Letting $P_n(t)$ be a polynomial of degree n, equation (3.7.28) can be written as

$$\int t^n \phi(t)\, dt = \phi(t)\, P_{n-1}(t) \text{ for n an odd integer} > 0. \qquad (3.7.29)$$

By using l'Hospital's Rule, it can be shown that

$$\lim_{t \to \infty} \phi(t)\, P_{n-1}(t) = \lim_{t \to \infty} \phi(t)\, P_{n-1}(t) = 0. \qquad (3.7.30)$$

Thus, using equations (3.7.29) and (3.7.30) to evaluate the limit yields

$$\lim_{\delta \to \infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t)\, dt = \lim_{\delta \to \infty} \left[\phi(\xi_\alpha-\delta)P_{n-1}(\xi_\alpha-\delta) - \phi(-\xi_\alpha-\delta)P_{n-1}(-\xi_\alpha-\delta) \right]$$

$$= 0$$

i.e.

$$\lim_{\delta \to \infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t)\, dt = 0 \text{ for n an odd integer} > 0. \qquad (3.7.31)$$

Consider now n to be an even, positive integer. Again, by repeated integration by parts,

$$\int t^n \phi(t)\, dt = -t^{n-1}\phi(t) - (n-1)t^{n-3}\phi(t) - \ldots$$
$$+ (n-1)(n-3) \ldots 1 \int \phi(t)\, dt \qquad (3.7.32)$$

Letting C(n) be a constant depending upon n, equation (3.7.32) reduces to

$$\int t^n \phi(t)\, dt = \phi(t) P_{n-1}(t) + C(n) \int \phi(t)\, dt$$

$$\text{for n an even integer} > 0. \qquad (3.7.33)$$

Now, defining $\Phi(x)$ to be the cumulative distribution function (cdf) of the N(0,1) distribution, i.e.

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-t^2/2} \, dt , \tag{3.7.34}$$

and noting by the property of cdf's that $\lim_{x \to -\infty} \Phi(x) = 0$, the limit becomes

$$\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t) \, dt = \lim_{\delta\to\infty} \begin{bmatrix} \phi(\xi_\alpha-\delta)P_{n-1}(\xi_\alpha-\delta) + C(n)\Phi(\xi_\alpha-\delta) \\ -\phi(-\xi_\alpha-\delta)P_{n-1}(\xi_\alpha-\delta) - C(n)\Phi(-\xi_\alpha-\delta) \end{bmatrix}$$

$$= 0.$$

Thus,

$$\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t) \, dt = 0 \text{ for } n \text{ an even integer} > 0. \tag{3.7.35}$$

For $n = 0$, $\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \phi(t) \, dt = 0$ by the property of the cdf mentioned above. This statement, along with equations (3.7.31) and (3.7.35), completes the proof to Lemma 3.3.

Lemma 3.4. Let m and n be non-negative integers and let ξ_α be a real, finite number. Then

$$\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta^m t^n \phi(t) \, dt = 0. \tag{3.7.36}$$

Proof: For $m = 0$ and n a non-negative integer, this is proved by Lemma 3.3. For $m > 0$ and n a non-negative integer, rewrite equation (3.7.36) as

$$\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta^m t^n \phi(t)\,dt = \lim_{\delta\to\infty} \frac{\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t)\,dt}{\delta^{-m}} . \tag{3.7.37}$$

The denominator has a limit of zero, and by Lemma 3.3 the numerator also has a limit of zero. Using l'Hospital's Rule on equation (3.7.37) yields

$$\lim_{\delta\to\infty} \frac{\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^n \phi(t)\,dt}{\delta^{-m}} = \lim_{\delta\to\infty} \frac{-(\xi_\alpha-\delta)^n \phi(\xi_\alpha-\delta) + (-\xi_\alpha-\delta)^n \phi(-\xi_\alpha-\delta)}{(-m)\,\delta^{-m-1}}$$

$$= m^{-1} \lim_{\delta\to\infty} \left[\delta^{m+1}(\xi_\alpha-\delta)^n \phi(\xi_\alpha-\delta) - \delta^{m+1}(-\xi_\alpha-\delta)^n \phi(-\xi_\alpha-\delta) \right]$$

$$= 0,$$

i.e.

$$\lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta^m t^n \phi(t)\,dt = 0 \qquad \text{for } m \text{ a positive integer, } n \text{ a non-negative integer.} \tag{3.7.38}$$

This completes the proof to Lemma 3.4.

Consider now $B(\delta)$ as $\delta \to \infty$. From equation (3.7.23)

$$\lim_{\delta\to\infty} B(\delta) = -w_2\sigma_z \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta\phi(t)\,dt$$

$$+ \frac{1}{\sigma_z} \left[-w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 k_2 m_x}{N_x} \right] \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t\phi(t)\,dt. \tag{3.7.39}$$

By Lemma 3.4 of this section, each term in equation (3.7.39) has a limit of zero, and thus

$$\lim_{\delta\to\infty} B(\delta) = 0. \tag{3.7.40}$$

H. Mean Square Error of the Two Stage, Preliminary Test Estimation Procedure

The mean square error (MSE) of the estimator $\hat{\mu}_y$ is defined to be

$$\mathrm{MSE}(\tilde{\hat{\mu}}_y) = E(\tilde{\hat{\mu}}_y - \mu_y)^2. \tag{3.8.1}$$

Expanding $(\tilde{\hat{\mu}}_y - \mu_y)^2$ yields

$$\mathrm{MSE}(\tilde{\hat{\mu}}_y) = E(\tilde{\hat{\mu}}_y)^2 - 2\,\mu_y\, E(\tilde{\hat{\mu}}_y) + \mu_y^2. \tag{3.8.2}$$

$\mathrm{MSE}(\tilde{\hat{\mu}}_y)$ will be derived by using equation (3.8.2). Since equation (3.7.21) gives $E(\tilde{\hat{\mu}}_y)$, it is necessary only to find $E(\tilde{\hat{\mu}}_y^2)$.

Following the general development in section F of this chapter where $E(\tilde{\hat{\mu}}_y)$ was derived, $E(\tilde{\hat{\mu}}_y^2)$ can be written as

$$\begin{aligned} E(\tilde{\hat{\mu}}_y^2) &= E(\tilde{v}^2 \mid |\tilde{z}| < \xi_\alpha \sigma_z)\Pr(|\tilde{z}| < \xi_\alpha \sigma_z) \\ &\quad + E(\tilde{\mu}^2 \mid |\tilde{z}| \geq \xi_\alpha \sigma_z)\Pr(|\tilde{z}| \geq \xi_\alpha \sigma_z) \end{aligned} \tag{3.8.3}$$

where $\tilde{u}$ and $\tilde{v}$ are the unpooled and pooled estimators derived in equations (3.3.6) and (3.3.5). By Lemma 3.2 and its corollary,

$$E(\tilde{\hat{\mu}}_y^2) = E(\tilde{\bar{y}}^2_{n+n_y+m_y}) + \int_{|z|<\xi_\alpha\sigma_z} E(\tilde{M} \mid z)\, h(z)\, dz \tag{3.8.4}$$

where $h(z)$ is defined in equation (3.7.2) and

$$M = (w_1 \bar{y}_{n+n_y} + w_2 \bar{x}_{n+n_x+m_x})^2 - \bar{y}^2_{n+n_y+m_y}\ . \tag{3.8.5}$$

Now, using the substitutions in equations (3.3.4) and expanding equation (3.8.5), M is obtained as

$$M = \bar{y}^2_{n+n_y} \left[\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\}^2 - \frac{(n+n_y)^2}{N_y^2} \right]$$

$$- 2\bar{y}_{n+n_y} \left[\begin{aligned} &\frac{w_2(n+n_x)}{N_x} \left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\} z + \frac{m_y(n+n_y)\bar{y}_{m_y}}{N_y^2} \\ &- \frac{w_2 m_x}{N_x} \left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\} \bar{x}_{m_x} \end{aligned} \right]$$

$$+ \frac{w_2^2(n+n_x)^2 z^2}{N_x^2} + \frac{w_2^2 m_x^2 \bar{x}^2_{m_x}}{N_x^2} - \frac{2w_2^2 m_x(n+n_x)z\bar{x}_{m_x}}{N_x^2}$$

$$- \frac{m_y^2 \bar{y}^2_{m_y}}{N_y^2} \tag{3.8.6}$$

Consider now the various conditional expectations in $E(\tilde{M} \mid z)$. First,

$$E(\tilde{\bar{y}}^2_{n+n_y} \mid z) = V(\tilde{\bar{y}}_{n+n_y} \mid z) + E^2(\tilde{\bar{y}}_{n+n_y} \mid z) \tag{3.8.7}$$

where $V(\tilde{\bar{y}}_{n+n_y} \mid z)$ and $E(\tilde{\bar{y}}_{n+n_y} \mid z)$ are given in equations (3.7.15) and 3.7.16). Also,

$$E(\tilde{z}\tilde{\bar{y}}_{n+n_y} \mid z) = z\, E(\tilde{\bar{y}}_{n+n_y} \mid z), \tag{3.8.8}$$

$$E(\tilde{\bar{x}}_{m_x} \tilde{\bar{y}}_{n+n_y} \mid z) = (\mu_y - \Delta)\, E(\tilde{\bar{y}}_{n+n_y} \mid z), \tag{3.8.9}$$

and

$$E(\tilde{\bar{y}}_{m_y}\tilde{\bar{y}}_{n+n_y}) = \mu_y\, E(\tilde{\bar{y}}_{n+n_y} \mid z). \tag{3.8.10}$$

Substituting the above expectations into $E(\tilde{M} \mid z)$ yields

$$E(\tilde{M} \mid z) = \left[\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\}^2 - \frac{(n+n_y)^2}{N_y^2} \right]\left[\left\{ \mu_y + \frac{k_1(z-\Delta)}{\sigma_z^2} \right\}^2 + \frac{\sigma_y^2}{n+n_y} - \frac{k_1^2}{\sigma_z^2} \right]$$

$$- 2\left\{ \mu_y + \frac{k_1(z-\Delta)}{\sigma_z^2} \right\} \left[\frac{w_2(n+n_x)}{N_x}\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\} z + \frac{m_y(n+n_y)\mu_y}{N_y^2} - \frac{w_2 m_x(\mu_y-\Delta)}{N_x}\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\} \right]$$

$$+ \frac{w_2^2(n+n_x)^2 z^2}{N_x^2} + \frac{w_2^2 m_x^2}{N_x^2}\left[(\mu_y-\Delta)^2 + \frac{\sigma_x^2}{m_x} \right] - \frac{m_y^2}{N_y^2}\left[\mu_y^2 + \frac{\sigma_y^2}{m_y} \right]$$

$$- \frac{2w_2^2 m_x(n+n_x)(\mu_y-\Delta)z}{N_x^2}. \tag{3.8.11}$$

If the terms in equation (3.8.11) are expanded, it can be shown that the coefficient of the quantity μ_y^2 is zero. Also, the coefficient of the quantity μ_y in equation (3.8.11) is

$$\frac{2k_1(z-\Delta)}{\sigma_z^2}\left[w_1 + \frac{w_2(n+n_x)}{N_x} - \frac{(n+n_y)}{N_y} \right] - 2w_2\left[\frac{(n+n_x)z}{N_x} + \frac{m_x\Delta}{N_x} \right]. \tag{3.8.12}$$

Hence, $E(\tilde{M} \mid z)$ can be written as

$$E(\tilde{M} \mid z) = 2\mu_y \left[\frac{k_1(z-\Delta)}{\sigma_z^2} \left\{ w_1 + \frac{w_2(n+n_x)}{N_x} - \frac{(n+n_y)}{N_y} \right\} - w_2 \left\{ \frac{(n+n_x)z}{N_x} + \frac{m_x \Delta}{N_x} \right\} \right]$$

$$+ \left[\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\}^2 - \frac{(n+n_y)^2}{N_y^2} \right] \left[\frac{k_1^2(z-\Delta)^2}{\sigma_z^4} + \frac{\sigma_y^2}{n+n_y} - \frac{k_1^2}{\sigma_z^2} \right]$$

$$- \frac{2k_1(z-\Delta)w_2}{\sigma_z^2} \left[\frac{z(n+n_x) + m_x \Delta}{N_x} \right] \left[w_1 + \frac{w_2(n+n_x)}{N_x} \right]$$

$$+ \frac{w_2^2(n+n_x)^2 z^2}{N_x^2} + \frac{w_2^2 m_x^2(\Delta^2+\sigma_x^2/m_x)}{N_x^2} + \frac{2w_2^2 m_x(n+n_x)\Delta\sigma_3}{N_x^2}$$

$$- \frac{m_y \sigma_y^2}{N_y^2} . \tag{3.8.13}$$

Now, using the change of variable $t = (z-\Delta)/\sigma_z$ in equation (3.8.4) yields

$$E(\tilde{\hat{\mu}}_y^2) = \mu_y^2 + \frac{\sigma_y^2}{N_y} + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} G(t)\,\phi(t)\,dt \tag{3.8.14}$$

where $G(t)$, as below, is obtained by taking $E(\tilde{M} \mid z)$ in equation (3.8.13), substituting $(t\sigma_z+\Delta)$ for z in $E(\tilde{M} \mid z)$, multiplying by the Jacabian of the transformation, and then expanding and collecting terms. Thus $G(t)$ is obtained as

$$G(t) = 2\mu_y \left[-w_2\Delta + \frac{t}{\sigma_z} \left\{ -w_2\sigma_z^2 + \frac{k_1 m_y}{N_y} + \frac{w_2 m_x k_2}{N_x} \right\} \right]$$

$$+ \frac{2tw_2\Delta}{\sigma_z} \left[w_2k_2 - w_1k_1 - \frac{w_2k_2m_x}{N_x} \right]$$

$$+ \frac{t^2}{\sigma_z^2} \left[\left\{ w_1k_1 - \frac{w_2k_2(n+n_x)}{N_x} \right\}^2 - \frac{k_1^2(n+n_y)^2}{N_y^2} \right]$$

$$+ \left[\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\}^2 - \frac{(n+n_y)^2}{N_y^2} \right] \left[\frac{\sigma_y^2}{n+n_y} - \frac{k_1^2}{k_1+k_2} \right]$$

$$+ \frac{w_2^2 m_x \sigma_x^2}{N_x^2} - \frac{m_y \sigma_y^2}{N_y^2} + w_2^{\,2}\Delta^2. \qquad (3.8.15)$$

With $E(\tilde{\hat{\mu}}_y^{\,2})$ as given in equation (3.8.14) and $E(\tilde{\hat{\mu}}_y)$ as in (3.7.21), substitute into equation (3.8.2) for $MSE(\tilde{\hat{\mu}}_y)$ and obtain $MSE(\tilde{\hat{\mu}}_y)$ as a function of δ, i.e. $MSE(\delta)$. Thus,

$$MSE(\delta) = \sigma_y^2/N_y + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (w_2^{\,2}\delta^2\sigma_z^2 + t^2A_2 + 2\delta tA_1 + A_0)\, \phi(t)\, dt \qquad (3.8.16)$$

where

$$A_2 = \sigma_z^{-2} \left[\left\{ w_1k_1 - \frac{w_2k_2(n+n_x)}{N_x} \right\}^2 - \frac{k_1^2(n+n_y)^2}{N_y^2} \right] \qquad (3.8.17)$$

$$A_1 = w_2 \left[w_2k_2 - w_1k_1 - w_2k_2m_x/N_x \right] \qquad (3.8.18)$$

$$A_0 = \left[\left\{ w_1 + \frac{w_2(n+n_x)}{N_x} \right\}^2 - \frac{(n+n_y)^2}{N_y^2} \right] \left[\frac{\sigma_y^2}{n+n_y} - \frac{k_1^2}{\sigma_z^2} \right] \tag{3.8.19}$$

$$+ \frac{w_2^2 m_x \sigma_x^2}{N_x^2} - \frac{m_y \sigma_y^2}{N_y^2} .$$

Note that MSE(δ) can be considered as a sum of two terms--the term σ_y^2/N_y, which is the MSE if one never pools and always takes N_y observations on the random variable $\tilde{y}$, and the term

$$\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (w_2^2\delta^2\sigma_z^2 + t^2A_2 + 2\delta tA_1 + A_0)\ \phi(t)\ dt$$

which can be either positive or negative.

Consider now MSE($-\delta$). From equation (3.8.16),

$$\text{MSE}(-\delta) = \sigma_y^2/N_y + \int_{-\xi_\alpha+\delta}^{\xi_\alpha+\delta} (w_2^2\delta^2\sigma_z^2 + t^2A_2 - 2\delta tA_1 + A_0)\ \phi(t)\ dt. \tag{3.8.20}$$

Using the change of variable $y = -t$, MSE($-\delta$) is obtained as

$$\text{MSE}(-\delta) = \sigma_y^2/N_y + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (w_2^2\delta^2\sigma_z^2 + y^2A_2 + 2\delta yA_1+A_0)\ \phi(y)\ dy,$$

i.e.

$$\text{MSE}(-\delta) = \text{MSE}(\delta). \tag{3.8.21}$$

Thus, it is necessary to investigate MSE(δ) for $\delta \geq 0$ only.

Consider now MSE(δ) as $\delta \longrightarrow \infty$. From equation (3.8.16),

$$\lim_{\delta\to\infty} \text{MSE}(\delta) = \sigma_y^2/N_y + A_2 \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t^2\phi(t)\,dt$$

$$+ w_2^2\sigma_z^2 \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta^2\phi(t)\,dt + 2A_1 \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \delta t\phi(t)\,dt$$

$$+ A_0 \lim_{\delta\to\infty} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \phi(t)\,dt. \qquad (3.8.22)$$

By Lemma 3.4 of this chapter, the limit of each term involving δ in equation (3.8.22) is zero, so that

$$\lim_{\delta\to\infty} \text{MSE}\ (\delta) = \sigma_y^2/N_y. \qquad (3.8.23)$$

This result confirms intuitive reasoning that as $\delta \to \infty$, the alternative hypothesis H_A: $\mu_y \neq \mu_x$ is always accepted and thus the unpooled estimator $\hat{\mu}_y = \bar{y}_{n+n_y+m_y}$, with variance σ_y^2/N_y, is always used.

I. Choice of the Weights w_1 and w_2

In the previous development w_1 and w_2 have been arbitrary constants such that they sum to one. Naturally, w_1 and w_2 should be selected advantageously. There are several possibilities.

If it is desired to minimize the variance of the pooled estimator, select w_1 and w_2 so as to minimize $V(w_1\bar{y}_{n+n_y} + w_2\bar{x}_{n+n_x+m_x})$. Since

$$\mathrm{cov}(\bar{y}_{n+n_y}, \bar{x}_{N_x}) = N_x^{-1}(n+n_y)^{-1} n\rho\sigma_x\sigma_y, \tag{3.9.1}$$

then

$$V(w_1\bar{y}_{n+n_y} + w_2\bar{x}_{N_x}) = \frac{\sigma_y^2(1-w_2)^2}{n+n_y} + \frac{\sigma_x^2 w_2^2}{N_x} + \frac{2w_2(1-w_2)n\rho\sigma_x\sigma_y}{N_x(n+n_y)}. \tag{3.9.2}$$

The solution which minimizes the above function is (w_1^*, w_2^*), where

$$w_1^* = \left[\frac{\sigma_x^2}{N_x} - \frac{n\rho\sigma_x\sigma_y}{N_x(n+n_y)}\right]\left[\frac{\sigma_y^2}{n+n_y} + \frac{\sigma_x^2}{N_x} - \frac{2n\rho\sigma_x\sigma_y}{N_x(n+n_y)}\right]^{-1} \tag{3.9.3}$$

and

$$w_2^* = \left[\frac{\sigma_y^2}{n+n_y} - \frac{n\rho\sigma_x\sigma_y}{N_x(n+n_y)}\right]\left[\frac{\sigma_y^2}{n+n_y} + \frac{\sigma_x^2}{N_x} - \frac{2n\rho\sigma_x\sigma_y}{N_x(n+n_y)}\right]^{-1}. \tag{3.9.4}$$

If now $m_x \doteq 0$, the weights w_1^* and w_2^* become

$$w_1^* \doteq \frac{k_2}{k_1+k_2}, \quad w_2^* \doteq \frac{k_1}{k_1+k_2}. \tag{3.9.5}$$

On the other hand, an investigation of MSE(δ) in equation (3.7.16) reveals that the term $\left[w_1k_1 - \frac{w_2k_2(n+n_x)}{N_x}\right]$ appears in both A_2 [equation (3.8.17)] and A_1 [equation (3.8.18)]. If it is desired to make this term zero, this yields another solution (w_1^+, w_2^+), where

$$w_1^+ = \frac{k_2(n+n_x)}{N_x}\left\{k_1 + \frac{k_2(n+n_x)}{N_x}\right\}^{-1} \tag{3.9.6}$$

and

$$w_2^+ = k_1 \left\{ k_1 + \frac{k_2(n+n_x)}{N_x} \right\}^{-1} . \tag{3.9.7}$$

Again, if $m_x \doteq 0$, this yields

$$w_1^+ \doteq \frac{k_2}{k_1+k_2} , \quad w_2^+ \doteq \frac{k_1}{k_1+k_2} . \tag{3.9.8}$$

A third approach may be to consider MSE(δ) in equation (3.8.16) as a function of w_2, and then minimize MSE(δ) with respect to w_2. This solution, however, leads to w_2 as a function of δ, which, of course, is unknown. This approach does not look very promising in the general case presented in this chapter, but it will be discussed later for some special cases.

IV. ONE STAGE PRELIMINARY TEST ESTIMATION SCHEME FOR THE MEAN OF A NORMAL POPULATION

A. Specification of Problem

This chapter discusses a special case of the general scheme in Chapter III by letting $m_x = m_y = n_x = n_y = 0$. Thus, the available sample data consitute a bivariate sample of size n. The preliminary test of H_o: $\mu_y = \mu_x$ versus H_A: $\mu_y \neq \mu_x$ uses the test statistic

$$z = \bar{y}_n - \bar{x}_n \tag{4.1.1}$$

with variance

$$V(\tilde{z}) = \sigma_z^2 = (\sigma_y^2 + \sigma_x^2 - 2\rho\sigma_x\sigma_y)/n. \tag{4.1.2}$$

The estimator $\hat{\mu}_y$ is defined as

$$\hat{\mu}_y = \begin{cases} \bar{y}_n & \text{if } |z| < \xi_\alpha\sigma_z \\ w_1\bar{y}_n + w_2\bar{x}_n & \text{if } |z| \geq \xi_\alpha\sigma_z \end{cases} \tag{4.1.3}$$

where $w_1 + w_2 = 1$ and ξ_α is defined in equations (3.7.3) and (3.7.4). This special case involves no element of survey planning, but would be applicable where the experimenter was presented with bivariate data and had no chance to do further sampling.

The main reason for considering this special case is to investigate the effect of the various parameters, especially ρ, on the bias and mean square error of the procedure. Kitagawa (1956, 1963) considered this problem when the parameters ρ and σ^2 are unknown, where $\sigma^2 = \sigma_x^2 = \sigma_y^2$. He showed that the mean square error and bias depend upon ρ through incomplete Beta functions, but he made no attempt to evaluate the effect of ρ on bias and mean square error. Assuming known covariance matrix, the formulas for bias and mean square error are less complicated than those of Kitagawa's, and a theoretical and numerical analysis provides some insight into the behavior of these two functions.

This chapter is organized in the following manner. First, some general properties of bias and mean square error are discussed in sections B and C when the weights w_1 and w_2 are arbitrary constants summing to one. The optimum weights are derived in section D, and there is some discussion about how to approximate them. In sections E and F some further properties of bias and mean square error are discussed when a particular set of weights is assumed. Section G gives some recommendations about the choice of α level for the preliminary test.

B. Bias with Arbitrary Weights w_1 and w_2

From equation (3.7.23), with $m_x = m_y = n_x = n_y = 0$, the bias of the preliminary test estimation procedure is obtained as

$$B(\delta) = -w_2 \sigma_z H(\delta), \tag{4.2.1}$$

where

$$H(\delta) = \int_{-\xi_\alpha - \delta}^{\xi_\alpha - \delta} (\delta + t)\, \phi(t)\, dt \tag{4.2.2}$$

and

$$k_1 = (\sigma_y^2 - \rho\sigma_x\sigma_y)/n$$

$$k_2 = (\sigma_x^2 - \rho\sigma_x\sigma_y)/n$$

$$k_1 + k_2 = \sigma_z^2 \qquad (4.2.3)$$

$$\delta = \Delta/\sigma_z = (\mu_y - \mu_x)/\sigma_z.$$

$B(\delta)$ can be calculated from tables of the $N(0, 1)$ density function $\phi(t)$ and the cumulative distribution function $\Phi(t)$ since

$$H(\delta) = \delta\left[\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right] - \phi(\xi_\alpha-\delta) + \phi(-\xi_\alpha-\delta). \qquad (4.2.4)$$

Note that $B(\delta)$ depends on δ only through $H(\delta)$ as long as w_2 is not a function of δ.

From the discussion in Chapter III, it can be concluded that

$$H(0) = 0$$

$$H(-\delta) = -H(\delta) \qquad (4.2.5)$$

$$\lim_{\delta\to\infty} H(\delta) = 0.$$

Lemma 4.1 gives an additional specification of $H(\delta)$.

Lemma 4.1. If $\xi_\alpha > 0$ is a real number, then $H(\delta) > 0$ for $\delta > 0$, where $H(\delta)$ is given in equation (4.2.2).

Proof: Writing $H(\delta)$ as the sum of two integrals over $(-\xi_\alpha-\delta, -\delta)$ and $(-\delta, \xi_\alpha-\delta)$, and using the change of variable $z = -(\delta + t)$ on the first integral and $w = (\delta + t)$ on the second integral, $H(\delta)$ is obtained as

$$H(\delta) = \int_0^{\xi_\alpha} t \left[\phi(t-\delta) - \rho(t+\delta)\right] \, dt. \tag{4.2.6}$$

Since

$$|t-\delta| < |t+\delta| \quad \text{for } t > 0 \text{ and } \delta > 0, \tag{4.2.7}$$

then

$$\phi(t-\delta) - \phi(t+\delta) > 0 \tag{4.2.8}$$

and hence

$$H(\delta) > 0 \text{ for } \delta > 0. \tag{4.2.9}$$

This completes the proof to Lemma 4.1.

With Lemma 4.1, equation (4.2.5) implies

$$H(\delta) < 0 \text{ for } \delta < 0. \tag{4.2.10}$$

Thus, for w_2 an arbitrary constant, $B(\delta)$ has the same algebraic sign for all $\delta > 0$ and the opposite algebraic sign for all $\delta < 0$. Usually $0 < w_2 < 1$, in which case $B(\delta) < 0$ for $\delta < 0$ and $B(\delta) > 0$ for $\delta < 0$. This seems intuitively correct, for $\delta > 0$ implies $\mu_y > \mu_x$, and thus a pooled estimator will underestimate μ_y.

From the numerical investigation cited later in this chapter, there is a strong indication that $H(\delta)$ has one critical point for $\delta > 0$ which gives a maximum value for $H(\delta)$. For the numerical investigation, $H(\delta)$ is represented in Figure 4.1 for $\delta \geq 0$ only, since $H(-\delta) = -H(\delta)$.

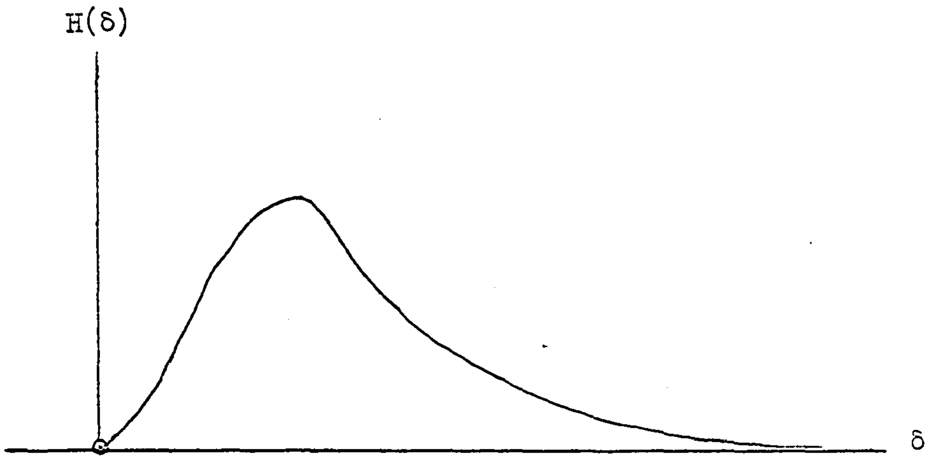

Figure 4.1. Form of $H(\delta)$ from numerical investigation

It is conjectured that this property is true for $H(\delta)$ in general.

The maximum value that $H(\delta)$ assumes for $\delta > 0$ depends only upon ξ_α, i.e. upon α. See Table 4.1 for the maximum value of $H(\delta)$ for various α levels and for what value of δ this maximum occurs. Once w_2 is specified, then the maximum possible bias, for any given α level, can be determined. Note that the maximum value of $H(\delta)$, as well as δ, increases as α decreases for the entries in Table 4.1. This is a result of another general property of $H(\delta)$ discussed in the following lemma.

Lemma 4.2. Let α_1 and α_2 be two sizes for the preliminary test, where $0 < \alpha_2 < \alpha_1 < 1$. Denote the function $H(\delta)$ with size α_1 as $H_{\alpha_1}(\delta)$, with the corresponding definition for $H_{\alpha_2}(\delta)$. Then $H_{\alpha_2}(\delta) > H_{\alpha_1}(\delta)$ for all $\delta > 0$.

Proof: Let the critical value $\xi_{\alpha_1} > 0$ correspond to a preliminary test of size α_1, with a similar definition for $\xi_{\alpha_2} > 0$. Then

$$\xi_{\alpha_2} = \xi_{\alpha_1} + \epsilon,\ \epsilon > 0. \tag{4.2.11}$$

By using equation (4.2.11) and the transformations $z = -(t+\xi_{\alpha_1}+\delta)$ and $w = (t-\xi_{\alpha_1}+\delta)$, $H_{\alpha_2}(\delta) - H_{\alpha_1}(\delta)$ can be written as

$$H_{\alpha_2}(\delta) - H_{\alpha_1}(\delta) = \int_0^{\epsilon} (z+\xi_{\alpha_1}) \left\{ \left[\phi(z+\xi_{\alpha_1}-\delta) - \phi(z+\xi_{\alpha_1}+\delta) \right] \right\} dz. \tag{4.2.12}$$

Since

$$|z + \xi_{\alpha_1} - \delta| < |z + \xi_{\alpha_1} + \delta| \text{ for } z > 0,\ \xi_{\alpha_1} > 0,\ \delta > 0, \tag{4.2.13}$$

then

$$\phi(z + \xi_{\alpha_1} - \delta) > \phi(z + \xi_{\alpha_1} + \delta) \tag{4.2.14}$$

and hence

$$H_{\alpha_2}(\delta) > H_{\alpha_1}(\delta) \text{ for } \delta > 0. \tag{4.2.15}$$

This concludes the proof to Lemma 4.2.

Lemma 4.2 implies that, for a fixed δ, $|B(\delta)|$ increases as α decreases. The choice of α which minimizes $|B(\delta)|$ is thus $\alpha = 1$. However, this precludes the possibility of ever pooling $\bar{x}_n$ and $\bar{y}_n$ and thus is not a satisfactory choice in this discussion. On the other hand, a choice of $\alpha = 0$ implies $B(\delta) = -w_2\delta\sigma_z$, which becomes infinite as $\delta \longrightarrow \infty$. Note that $B(\delta) \longrightarrow 0$ as $\delta \longrightarrow \infty$ if, and only if, ξ_α is finite, i.e. $\alpha > 0$. Hence, a choice of α near zero will allow a great chance of pooling but can also admit a substantial bias, whereas a choice of α near one will minimize bias but will allow only a small chance of pooling. Some compromise choice of α which will allow both pooling and moderate bias is probably the best strategy.

A further description of $H(\delta)$ for various α levels is given in Table 4.2. For all α levels except $\alpha = .01$, $H(\delta)$ reaches a maximum value for $1.0 < \delta < 1.5$ and decreases almost to zero around $\delta = 3.46$.

C. Mean Square Error with Arbitrary Weights w_1 and w_2

From equation (3.8.16) in Chapter III, with the modifications discussed in section B of this chapter, the mean square error is obtained as

$$MSE(\delta) = \sigma_y^2/n + K(\delta) \tag{4.3.1}$$

where

$$K(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left\{ \begin{array}{l} w_2^2\delta^2\sigma_z^2 + 2\delta t w_2(w_2k_2-w_1k_1) \\ -w_2t^2(w_1k_1-w_2k_2+k_1) \end{array} \right\} \phi(t)\, dt. \tag{4.3.2}$$

From the discussion in Chapter III it can be concluded that

$$\begin{aligned} &K(-\delta) = K(\delta) \\ &\lim_{\delta\to\infty} K(\delta) = 0. \end{aligned} \tag{4.3.3}$$

It seems reasonable to expect that $K(0) < 0$, thus giving a smaller mean square error for the sometimes pool procedure when $\delta = 0$ than for the never pool procedure. From equation (4.3.2),

$$K(0) = -w_2(w_1k_1-w_2k_2+k_1) \int_{-\xi_\alpha}^{\xi_\alpha} t^2\phi(t)\, dt. \tag{4.3.4}$$

Until w_1 and w_2 are specified, however, it is not certain that $K(0) < 0$. For example, if $w_1 = w_2 = 1/2$, $\sigma_y^2 = 1$, $\sigma_y^2 = 6$, and $\rho = 0$, then $K(0) > 0$. This result indicates that $w_1 = w_2 = 1/2$ is probably a poor choice of weights in this particular example.

No meaningful general analysis can be done on $K(\delta)$ until the weights w_1 and w_2 are further specified.

D. The Optimum Weights w_{10} and w_{20}

Consider now $K(\delta)$ in equation (4.3.2) to be a function of w_2, i.e. $K(\delta) = \phi(w_2)$. Substituting $(1-w_2)$ for w_1 and collecting terms of like powers in w_2 yields

$$\begin{aligned}\phi(w_2) &= w_2^2(k_1+k_2)\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}(\delta+t)^2\,\phi(t)\,dt\\ &\quad - 2w_2k_1\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}(t^2+\delta t)\,\phi(t)\,dt.\end{aligned} \tag{4.4.1}$$

To minimize $\phi(w_2)$ with respect to w_2, solve $\phi'(w_{20}) = 0$ for w_{20} (optimum value of w_2). This yields

$$w_{20} = k_1A(\delta)\Big/\Big\{(k_1+k_2)C(\delta)\Big\} \tag{4.4.2}$$

where

$$A(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}(t^2+\delta t)\,\phi(t)\,dt \tag{4.4.3}$$

and

$$C(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (\delta+t)^2 \phi(t)\, dt. \qquad (4.4.4)$$

Note that $A(-\delta) = A(\delta)$ and $C(-\delta) = C(\delta)$. Thus, the same optimum weight w_{20} is obtained for $\delta = c$ and $\delta = -c$, c a real number.

Substituting w_{20} into $\phi(w_2)$ in equation (4.4.1) yields

$$\phi(w_{20}) = -k_1^2 A^2(\delta) / \left[(k_1+k_2) C(\delta) \right] . \qquad (4.4.5)$$

Since $C(\delta) > 0$ and $(k_1+k_2) > 0$, then $\phi(w_{20}) < 0$ for all δ. Thus, if $w_2 = w_{20}$ and $w_1 = 1-w_{20}$, then $MSE(\delta) < \sigma_y^2/n$ for all δ. Hence, there exist weights w_2 and w_1 such that the mean square error of the preliminary test procedure is always less than the mean square error of the never pool procedure.

Using Lemma 3.4 and l'Hospital's Rule, it can be shown that

$$A(\delta)/C(\delta) \longrightarrow -\infty \quad \text{as } \delta \longrightarrow \infty. \qquad (4.4.6)$$

Thus, the optimum weight w_{20} approaches $-\infty$ if $k_1 > 0$ and $+\infty$ if $k_1 < 0$.

Consider now the bias and mean square error when the optimum weight w_{20} is used. From equation (4.2.1), the bias when $w_2 = w_{20}$, denoted by $B(w_{20})$, is

$$B(w_{20}) = -k_1 H(\delta) A(\delta) / \left[\sigma_z C(\delta) \right] . \qquad (4.4.7)$$

The properties of $B(\delta)$ enumerated in equation (4.2.5) for a constant weight w_2 also hold when $w_2 = w_{20}$, i.e.

$$B(w_{20}) = 0 \text{ when } \delta = 0$$

$$B(w_{20}) \text{ is antisymmetric in } \delta \tag{4.4.8}$$

$$\lim_{\delta \to \infty} B(w_{20}) = 0.$$

Similarly, the mean square error when $w_2 = w_{20}$ is

$$MSE(w_{20}) = \sigma_y^2/n + \phi(w_{20}), \tag{4.4.9}$$

where $\phi(w_{20})$ is given in equation (4.4.5). Also, it can be shown that

$$\lim_{\delta \to \infty} MSE(w_{20}) = \sigma_y^2/n. \tag{4.4.10}$$

Thus, even though the optimum weight w_{20} becomes infinite as $\delta \longrightarrow \infty$, the bias and mean square error with this optimum weight approach the limits zero and σ_y^2/n, respectively.

Now, of course, the optimum weights cannot be used directly since Δ, and thus δ, is unknown. Some other possibilities exist, however. First, if a good guess of δ exists, then the optimum weights with δ replaced by the guessed value of δ may be useful. Secondly, the replacement of δ by $\hat{\delta}$, where $\hat{\delta}$ is an estimate of δ provided by the sample, may result in useful weights. This procedure is investigated by Monte Carlo methods in the next chapter. A third approximation combines the classical and Bayesian approach by assuming a prior distribution $g(\delta)$ on $\tilde{\delta}$. Then $K(\delta)$, as given in equation (4.3.2), can be averaged over $g(\delta)$, which yields $E[K(\tilde{\delta})]$ as a function of w_2, k_1, and k_2. w_2 can then be chosen so as to minimize $E[K(\tilde{\delta})]$ with respect to w_2. Thus, w_2 will be a function of k_1 and k_2, which are known. Alternatively, the function w_{20} in equation (4.4.2) can be averaged over $g(\delta)$ to yield a weight which is also a function of k_1 and k_2. These two procedures will probably not produce the same weight. Note, however, that these expectations are not trivial since δ is involved as a limit of integration in both $K(\delta)$ and w_{20}. Expansion of $K(\delta)$ and w_{20} in an infinite series may aid in evaluating $E[K(\tilde{\delta})]$ and $E(\tilde{w}_{20})$.

In ignorance of the value of δ, it seems best to evaluate w_{20} at $\delta = 0$, which then yields

$$\begin{aligned} w_2 &= k_1/(k_1+k_2) \\ w_1 &= k_2/(k_1+k_2). \end{aligned} \tag{4.4.11}$$

This solution is advocated because the weights in equation (4.4.11) are also the weights which minimize the variance of the pooled estimator $w_1\bar{y}_n + w_2\bar{x}_n$.

Note that the weights $k_1/(k_1+k_2)$ and $k_2/(k_1+k_2)$ may be negative or larger than one. A necessary and sufficient condition that both weights be between zero and one is that $\rho < \min\left[\sigma_x/\sigma_y,\ \sigma_y/\sigma_x\right]$. Of course, if $\rho < 0$ this condition is satisfied for any σ_x^2 and σ_y^2. If there is a large difference in the magnitude of σ_x^2 and σ_y^2, then $\min\left[\sigma_x/\sigma_y,\ \sigma_y/\sigma_x\right]$ is small, and a large correlation will result in one negative weight and one weight larger than one. For example, if $\sigma_y^2 = 1$ and $\sigma_x^2 = 9$, then w_1 and w_2 are between zero and one if, and only if, $\rho < 1/3$. Since $w_1\bar{y}_n+w_2\bar{x}_n$ can be written as $\bar{x}_n+w_1(\bar{y}_n-\bar{x}_n)$, the weight w_1 can be regarded as a correction factor to the estimator $\bar{x}_n$ and thus could easily be negative or larger than one.

E. Mean Square Error and Bias when $w_2=k_1/(k_1+k_2)$

Substituting $w_2=k_1/(k_1+k_2)$ into equations (4.3.1) and (4.3.2) yields

$$MSE(\delta) = \sigma_y^2/n + k_1^2 \psi(\delta)/(k_1+k_2) \tag{4.5.1}$$

where

$$\psi(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (\delta^2-t^2)\, \phi(t)\, dt. \tag{4.5.2}$$

$MSE(\delta)$ can be calculated with a table of the $N(0, 1)$ density function $\phi(t)$ and the cumulative distribution function $\Phi(t)$ since

$$\begin{aligned}\psi(\delta) = (\delta^2-1) &\left\{\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right\} \\ &+ (\xi_\alpha-\delta)\, \phi(\xi_\alpha-\delta) - (-\xi_\alpha-\delta)\, \phi(-\xi_\alpha-\delta).\end{aligned} \tag{4.5.3}$$

From equation (4.3.3),

$$\begin{aligned}&\psi(-\delta) = \psi(\delta) \\ &\lim_{\delta\to\infty} \psi(\delta) = 0.\end{aligned} \tag{4.5.4}$$

Evaluating $\psi(\delta)$ at $\delta = 0$ yields

$$\psi(0) = -\int_{-\xi_\alpha}^{\xi_\alpha} t^2\, \phi(t)\, dt < 0. \tag{4.5.5}$$

Thus, the mean square error for the sometimes pool procedure is smaller, when $\mu_y = \mu_x$, than the mean square error for the never pool procedure. By considering the first and second derivative of $\psi(\delta)$ with respect to δ, it can be shown that $\delta = 0$ gives a local minimum for $\psi(\delta)$, which is negative. The minimum value of $\psi(\delta)$ for various α levels is given in Table 4.3. From numerical results cited later in this chapter, there is strong indication that for $\delta \geq 0$ the function $\psi(\delta)$ has an absolute mimimum for $\delta = 0$, increases to a positive absolute maximum for $\delta > 0$ and then decreases to zero asymptotically. Figure 4.2 illustrates the general shape of $\psi(\delta)$ for the numerical results for $\delta \geq 0$ only, since $\psi(-\delta) = \psi(\delta)$. It is conjectured that this property is true for $\psi(\delta)$ in general.

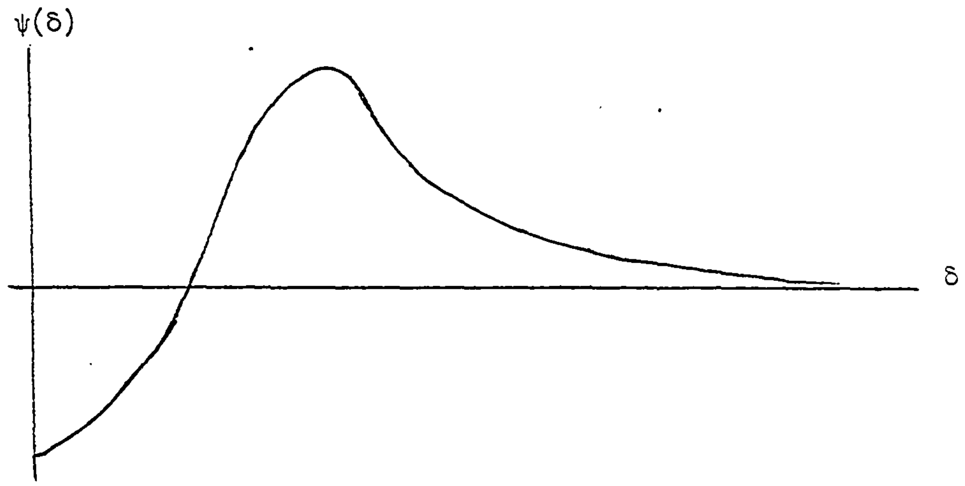

Figure 4.2. Form of $\psi(\delta)$ from numerical investigation

The size α of the preliminary test effects MSE(δ) only through $\psi(\delta)$. Table 4.3 illustrates that a decrease in α results in $\psi(\delta)$ having a smaller minimum value and a larger maximum value. Thus, for small α, a substantial decrease in mean square error over σ_y^2/n can be obtained if δ is near zero, but at the same time there is the risk of a large increase in mean square error over σ_y^2/n if δ is in the range (1.6, 2.6). Note from Table 4.4 that $\psi(\delta)$ is almost zero asymptotically by the time δ is between four and six.

When $w_2=k_1/(k_1+k_2)$, equation (4.2.1) gives the bias as

$$B(\delta) = -k_1 H(\delta)/\sigma_z \ . \tag{4.5.6}$$

F. Numerical Investigation

1. Parameter values for numerical work

Any further theoretical anlysis of $B(\delta)$ and $MSE(\delta)$ is not rewarding because it is not clear how various factors interact. For this reason, $B(\delta)$ and $MSE(\delta)$, as given in equations (4.5.6) and (4.5.1), were evaluated on the IBM 360/50 computer for various value of the parameters. Several combinations (not all) of the following parameter values were considered.

α = .10, .25, .50

n = 4, 9, 15, 25

Δ = 0, .5, 1.0, 1.5, 2.0, 3.0, 4.0

ρ = -2/3, -1/3, 0, 1/3, 1/2, 2/3, 3/4, 7/8

(σ_y^2, σ_x^2) = (9,9), (9,15), (9,18), (6,4), (6,6), (6,10), (6,12), (6,18), (6,24), (3,2)

Tables 4.5 through 4.10 give the value of δ, $B(\delta)$, and $MSE(\delta)$ for n = 15; α = .50, .25, .10; (σ_y^2, σ_x^2) = (6,24), (6,12), and various values of ρ and Δ. These six tables enable the investigation of the effect of variation in each of the parameters ρ, Δ, α, and (σ_y^2, σ_x^2) when all other parameters are fixed. The remaining tables from the numerical evaluation are not presented since they illustrate the same points brought out by these six tables.

2. Investigation of bias

Figures 4.5 and 4.6 graph all of the information about bias contained in Table 4.5 except the bias entries for $\rho = 3/4$ and $\rho = 7/8$. In Figure 4.5, $-B(\delta)$ is graphed against Δ, whereas in Figure 4.6 $-B(\delta)$ is graphed against δ. The relationship between the graphs in the two figures is very similar. In Figure 4.6 all of the graphs reach a maximum at the same value of δ, and from Table 4.1 this value of δ should be 1.05. In Figure 4.5 note that as ρ increases the maximum value of $-B(\delta)$ is obtained for a smaller value of Δ. This is because δ is an increasing function of ρ, and for large ρ, $\delta = 1.05$ is obtained for a smaller value of Δ. The general shape of the graphs is due to the function $H(\delta)$ as represented in Figure 4.1, since

$$B(\delta) = -k_1 H(\delta)/\sigma_z . \qquad (4.6.1)$$

Figures 4.5 and 4.6 illustrate that an increase in ρ will decrease $|B(\delta)|$ when all other parameters remain fixed. This is not universally true, however. Consider the column for $\Delta = 0.5$ in Table 4.5 and read the entries of $B(\delta)$ as ρ increases from $-2/3$ to $7/8$. Note that $B(\delta)$ starts at -0.0134, approaches zero, changes algebraic sign somewhere between $\rho = 2/3$ and $\rho = 3/4$, and then starts increasing. On the other hand, if the column $\Delta = 3.0$ in Table 4.5 is considered, then $B(\delta)$ starts at -0.0104 for

$\rho = -2/3$, approaches zero as ρ increases, and remains at zero for further increases in ρ. Behavior of $B(\delta)$ similar to this can be noted in all six tables whenever ρ increases and all other parameters remain fixed.

The reason for the change of algebraic sign of $B(\delta)$ for variation in ρ is easy to discern. Since $H(\delta) > 0$ for $\delta > 0$ and all values of δ considered in Tables 4.5 through 4.10 are positive, then the sign of $B(\delta)$ is determined by the sign of $-k_1$. From equation (4.2.3) $k_1 > 0$ if, and only if, $\rho < \sigma_y/\sigma_x$. Thus, in Tables 4.5 through 4.7, $B(\delta)$ is negative for $\rho < \sqrt{2}/2$, and positive for $\rho > \sqrt{2}/2$. In Tables 4.8 through 4.10 $B(\delta)$ is negative for $\rho < 1/2$ and positive for $\rho > 1/2$. Note in Tables 4.8 through 4.10 that $B(\delta) = 0$ for all δ and Δ when $\rho < 1/2$. In general, $B(\delta) = 0$ for $\rho = \sigma_y/\sigma_x$, since $\rho\ \sigma_y/\sigma_x$ implies $k_1 = 0$.

An explanation of the remaining behavior of $B(\delta)$ for variation in ρ can be obtained by considering $B(\delta)$ as a multiplicative function of the two factors $-k_1/\sigma_z$ and $H(\delta)$. The behavior of $H(\delta)$ for changes in δ was studied in detail in section B of this chapter. With all parameters except ρ fixed, k_1/σ_z can be considered as a function of ρ, and Figure 4.3 illustrates k_1/σ_z graphically for $-1 \leq \rho \leq 1$.

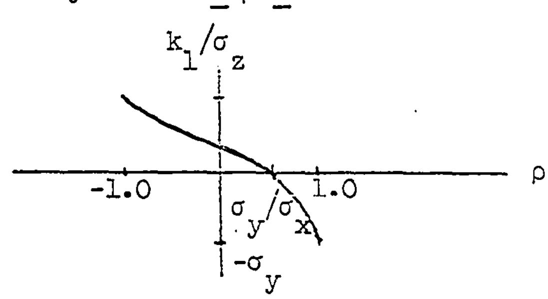

Figure 4.3. k_1/σ_z as a function of ρ

Thus, the coefficient $-k_1/\sigma_z$ causes $B(\delta)$, $\delta > 0$, to approach zero from below as $\rho \longrightarrow \sigma_y/\sigma_x$. For $\rho > \sigma_y/\sigma_x$, $-k_1/\sigma_z$ is positive and causes $B(\delta)$ to increase. However, as ρ increases δ also increases, and for δ sufficiently large (e.g. $\delta = 3$ in Table 4.5) $H(\delta)$ causes $B(\delta)$ to approach zero irregardless of the value of $-k_1/\sigma_z$ since $H(\delta) \longrightarrow 0$ as δ increases beyond 3. For an example of the dominant behavior of $H(\delta)$, see the column $\Delta = 3.0$ in Table 4.5 for the change in $B(\delta)$ as ρ increases.

Figure 4.7 illustrates the effect on $B(\delta)$ of variation in α. Of course, the results graphed in Figure 4.6 also follow directly from Lemma 4.2.

Figure 4.8 illustrates the effect on $B(\delta)$ of changing (σ_y^2, σ_x^2) from (6,12) to (6,24). Note that $-B(\delta)$ is larger for $(\sigma_y^2, \sigma_x^2) = (6,12)$ than for $(\sigma_y^2, \sigma_x^2) = (6,24)$ for $\Delta < 1.8$, but for $\Delta > 1.8$, $-B(\delta)$ is larger for $(\sigma_y^2, \sigma_x^2) = (6,24)$. Again, a consideration of $B(\delta)$ as a multiplicative function of the two factors $-k_1/\sigma_z$ and $H(\delta)$ can be helpful in explaining this behavior. It can be shown that k_1/σ_z is a decreasing function of σ_x for $\sigma_x > 0$ and $-1 < \rho < 1$. Thus, $\sigma_x^2 = 24$ yields a smaller value of k_1/σ_z than $\sigma_x^2 = 12$, and for $\Delta < 1.8$ in Figure 4.8, a smaller value of $-B(\delta)$ is obtained for $\sigma_x^2 = 24$. However, $\sigma_x^2 = 24$ yields a smaller value of δ than $\sigma_x^2 = 12$, and for $\Delta > 1.8$ in Figure 4.8, $H(\delta)$ becomes the dominant multiplicative factor in $-B(\delta)$. Thus, $-B(\delta)$ is larger for $\sigma_x^2 = 24$ when $\Delta > 1.8$ in Figure 4.8.

3. Investigation of mean square error

Figures 4.9 and 4.10 graph most of the information about MSE(δ) contained in Table 4.5. In Figure 4.9, MSE(δ) is graphed against Δ, whereas in Figure 4.10 MSE(δ) is graphed against δ. In Figure 4.10, which is for $\alpha = .50$, all graphs cross the line MSE(δ) $= .4000 = \sigma_y^2/n$ at $\delta = .72$. and reach a maximum at $\delta = 1.65$. This is due to the influence of $\psi(\delta)$, described for various α levels in Tables 4.3 and 4.4 and illustrated in Figure 4.2.

Equations (4.5.1) and (4.5.2) illustrate that Δ or δ effect MSE(δ) only through the function $\psi(\delta)$. From Table 4.4 MSE(δ) $< \sigma_y^2/n$ for $\delta < .72$ if $\alpha = .50$, for $\delta < .75$ if $\alpha = .25$, and for $\delta < .80$ if $\alpha = .10$. Thus, for values of δ around 3/4 or less, the sometimes pool procedure provides an increase in precision over the never pool procedure. For example, if $\sigma_x^2 = \sigma_y^2 = \sigma^2$, $\rho = 0$, and $n = 18$, then $\delta < 3/4$ if, and only if, $\Delta/\sigma < 1/4$. MSE(δ) reaches its maximum value at $\delta = 1.6$ for $\alpha = .50$, at $\delta = 1.8$ for $\alpha = .25$, and $\delta = 2.0$ for $\alpha = .10$. Hence, for moderate values of δ between 1.5 and 2.0, the precision of the sometimes pool procedure is much less than the precision of the never pool procedure. For δ larger than 4.0, there appears to be relatively small loss in precision due to using the sometimes pool procedure with α levels of .50, .25, and .10.

From Figures 4.9 and 4.10 it appears that an increase in ρ causes more variation in MSE(δ); i.e. if ρ is increased, then MSE(δ) graphed as a function of either δ or Δ will have a lower minimum and higher maximum. This is not always true. For example, if the MSE(δ) entries in Table 4.5 are read for $\rho = 3/4$, it can be seen that MSE(δ) has less variation for $\rho = 3/4$ than for $\rho = 1/2$. Recall from equations (4.5.1) and (4.5.2) that

$$\mathrm{MSE}(\delta) = \sigma_y^2/n + k_1^2\ \psi(\delta)/(k_1+k_2) \tag{4.6.2}$$

and

$$\psi(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (\delta^2-t^2)\ \phi\ (t)\ dt. \tag{4.6.3}$$

It will be helpful in explaining the behavior of MSE(δ) to consider the quantity $k_1^2\psi(\delta)/(k_1+k_2)$ as a multiplicative function of the two terms $k_1^2/(k_1+k_2)$ and $\psi(\delta)$.

The term $\psi(\delta)$ was discussed in section E of this chapter. From Table 4.4, $\psi(\delta) \doteq 0$ for $\delta \doteq .75$ and $\delta \in (4,5)$ for $\alpha = .50, .25, .10$.

Consider now the term $k_1^2/(k_1+k_2)$. Figure 4.4 graphs $k_1^2/(k_1+k_2)$ as a function of ρ for $-1 \leq \rho \leq 1$ and $\sigma_x^2 > \sigma_y^2$.

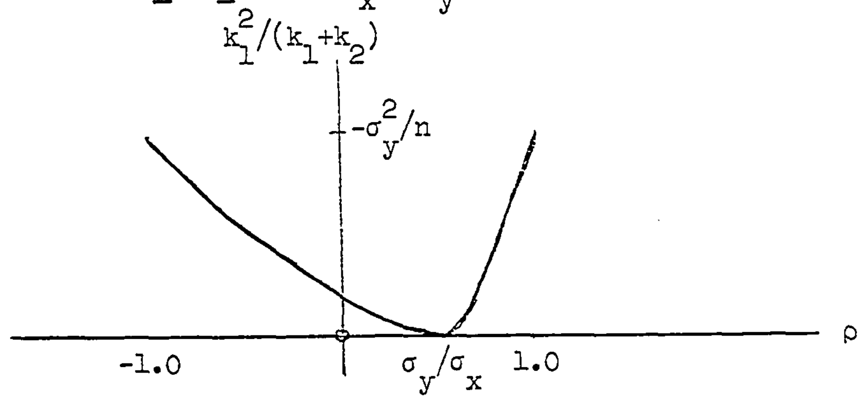

Figure 4.4. $k_1^2/(k_1+k_2)$ as a function of ρ

Thus, if $\rho \doteq \sigma_y/\sigma_x$, MSE($\delta$) will be approximately equal to σ_y^2/n.

If now ρ is varied while all other parameters are held fixed, the variation in MSE(δ) can be accounted for by noting how $k_1^2/(k_1+k_2)$ and $\psi(\delta)$ behave as ρ changes. For example, in the column $\Delta = 0$ in Table 4.5, note that MSE(δ) begins at 0.3780, increases to 0.4000 at $\rho \doteq .7$, and then decreases again. Since $\psi(\delta) < 0$ for $\delta = 0$ and is also constant with respect to ρ, then MSE(δ) increases as $\rho \longrightarrow \sigma_y/\sigma_x$ because $k_1^2/(k_1+k_2)$ is decreasing. Similarly, for $\rho > \sigma_y/\sigma_x$ MSE(δ) decreases because $k_1^2/(k_1+k_2)$ is increasing. Note that in the column $\Delta = 0.5$ in Table 4.5, the entries for MSE(δ) are increasing. $\psi(\delta)$ has the main effect here since it is an increasing function of δ for $\delta < 1.6$ and all values of δ in the column $\Delta = 0.5$ are less than 1.1. Note also that MSE(δ) $\doteq \sigma_y^2/n$ for $\rho \doteq \sigma_y/\sigma_x$ since then $k_1^2/(k_1+k_2) \doteq 0$. The remaining tables can be analyzed similarly for variation in MSE(δ) caused by variation in ρ. Such an analysis must always consider how ρ effects $k_1^2/(k_1+k_2)$ and what value of δ the particular value of ρ is producing.

Figure 4.11 illustrates the effect on MSE(δ) of variation in α and Δ (i.e. δ) when all other parameters are fixed. A higher value of α gives a more conservative procedure, i.e. a mean square error with less difference between its minimum and maximum values.

Figure 4.12 illustrates the effect on MSE(δ) of changing (σ_y^2, σ_x^2) from (6,12) to (6,24). When $(\sigma_y^2, \sigma_x^2) = (6,24)$, a more conservative procedure results, i.e. less possible gain but also less possible loss in precision.

G. Recommendations on Choice of α Level

When a decision must be made whether to use a sometimes pool procedure, and if so, what α level to use, control of MSE(δ) is usually of more interest that B(δ). Tables 4.3 and 4.4 give some guidelines for choice of α level.

For example, if it is known that $\delta < .905$, the sometimes pool procedure can be used at $\alpha = .01$ and the resulting mean square error will be less than σ_y^2/n. If $\delta < .75$, then a choice of $\alpha = .25$ would insure the same thing.

Recall from equation (4.6.2) that

$$\mathrm{MSE}(\delta) = \sigma_y^2/n + k_1^2\psi(\delta)/(k_1+k_2) \tag{4.7.1}$$

If it is known that $\delta > 4.77$, then the resultant mean square error will be less than $\sigma_y^2/n + k_1^2(.01)/(k_1+k_2)$ for $\alpha = .10$. If the increase over σ_y^2 / n is too high, then this increase can be reduced by selecting a larger value of α. Suppose now that the experimenter is willing to tolerate a 10% possible increase in mean square error over σ_y^2/n by using the preliminary test procedure. The increase over σ_y^2/n is $k_1^2\psi(\delta)/(k_1+k_2)$ and thus the experimenter's criterion is

$$\frac{k_1^2\psi\,(\delta)n}{(k_1+k_2)\sigma_y^2} \leq .10. \tag{4.7.2}$$

For example, if $\sigma_y^2 = 4$, $\sigma_x^2 = 9$, and $\rho = 1/3$, then $k_1 = 2/n$, $\sigma_x^2 = 9/n$, and inequality (4.7.2) becomes

$$\psi(\delta) \leq .9. \tag{4.7.3}$$

If the experimenter wishes to use $\alpha = .25$, then Table 4.3 indicates that inequality (4.7.3) will be satisfied for all values of δ since $\max_{\delta} \psi(\delta) = 0.379$ for $\alpha = .25$. Likewise, a choice of $\alpha = .50$ will satisfy inequality (4.7.3), and a choice of $\alpha = .10$ nearly satisfies (4.7.3). If, however, the experimenter wants to use $\alpha = .01$, then his 10% criterion is not satisfied for those values of δ (approximately) such that

$$1.3 < \delta < 4.0 \tag{4.7.4}$$

since Table 4.4 indicates that for $1.366 < \delta < 3.967$, $\psi(\delta) \geq 1.0$. Now, $\delta = \Delta\sqrt{n}/3$, and thus inequality (4.7.4) becomes (approximately)

$$4/\sqrt{n} < \Delta < 12/\sqrt{n} \,. \tag{4.7.5}$$

Thus, if the experimenter wants to use $\alpha = .01$ and satisfy his 10% criterion, he must be convinced that Δ satisfies inequality (4.7.5) for his choice of sample size n. If this is unreasonable or if the experimenter has no idea at all about the size of Δ, then it is best in this example to use $\alpha = .50$, .25, or .10 where the 10% criterion is then satisfied approximately for any value of δ or Δ. If a selection of at least $\alpha = .50$ will not give the required precision, then it is probably best not to use the preliminary test procedure, since money can then be saved by not measuring any observations on the random variable $\tilde{x}$.

Tables 4.1 and 4.2 give similar values for the function $H(\delta)$. For example, if $|\delta| < .38$ and $\alpha = .25$, then $|B(\delta)| < |k_1|(.1)/\sigma_z$. On the other hand, if $|\delta| > 2.15$ and $\alpha = .25$, then $|B(\delta)| < |k_1|(.1)/\sigma_z$. With Tables 4.1 and 4.2 and a reasonable knowledge of δ, an α level can be selected so that $|B(\delta)|$ will be less than some preassigned constant. If the experimenter can tolerate, for example, a bias which is 5% of the non-pool standard error $\sigma_y/\sqrt{n}$, then an α level can be selected using Tables 4.1 and 4.2 by a method similar to the one used in the example about mean square error.

If nothing is known about δ, then the maximum value for $B(\delta)$ and $MSE(\delta)$ for given covariance matrix and α level can be obtained from Tables 4.1 and 4.3. If the maximum possible bias and mean square error cannot be tolerated, then in the absence of any knowledge about δ it seems best not to use the preliminary test procedure.

H. Summary

From Figures 4.5 through 4.12 and from the discussion of this chapter, some general conclusions about $B(\delta)$ and $MSE(\delta)$ as functions of α, ρ, and σ_x^2 can be reached. First, as α increases, $|B(\delta)|$ decreases. Also, as α increases, $MSE(\delta)$ is dampened, i.e. it has a larger minimum value and a smaller maximum value.

The following conclusions hold most of the time, but not all of the time. Counterexamples are pointed out in section F. First, as ρ increases, $|B(\delta)|$ decreases. Also, as ρ increases, $MSE(\delta)$ is dampened as mentioned in the preceding paragraph. Second, as σ_x^2 increases, $MSE(\delta)$ is dampened.

Also, as σ_x^2 increases, $B(\delta)$ has a smaller maximum value, but approaches zero more slowly.

Table 4.1. δ and $H(\delta)$ such that $H(\delta)$ is maximum

α	δ	$H(\delta)$
.50	1.047	.045
.25	1.144	.192
.10	1.312	.453
.05	1.459	.658
.01	1.824	1.111

Table 4.2. Values of δ such that $H(\delta) = c$ for given c and α

c	α=.50	α=.25	α=.10	α=.05	α=.01
0.001	0.014	0.004	0.002	0.002	0.002
0.01	0.142	0.037	0.018	0.014	0.011
0.10	none	0.383	0.180	0.140	0.110
1.00	none	none	none	none	1.319
$\max_{\delta} H(\delta)$	1.047	1.144	1.312	1.459	1.824
1.00	none	none	none	none	2.341
0.10	none	2.149	3.023	3.470	4.259
0.01	2.512	3.394	3.394	3.394	5.195
0.001	3.466	3.466	3.466	3.466	5.907

Table 4.3. δ and $\psi(\delta)$ such that $\psi(\delta)$ is a minimum and maximum

	minimum		maximum	
α	δ	$\psi(\delta)$	δ	$\psi(\delta)$
.50	0.0	-0.071	1.648	0.087
.25	0.0	-0.276	1.780	0.379
.10	0.0	-0.561	1.993	0.952
.05	0.0	-0.721	2.164	1.464
.01	0.0	-0.916	2.554	2.819

Table 4.4. Values of δ such that $\psi(\delta)$ = c for given c and α

c	α=.50	α=.25	α=.10	α=.05	α=.01
0.0	0.724	0.754	0.801	0.835	0.905
0.001	0.730	0.756	0.802	0.837	0.908
0.01	0.788	0.771	0.810	0.843	0.913
0.10	none	0.923	0.885	0.902	0.959
1.0	none	none	none	1.489	1.366
$\max_\delta \psi(\delta)$	1.648	1.780	1.993	2.164	2.554
1.0	none	none	none	2.949	3.967
0.10	none	3.123	3.903	4.334	5.114
0.01	3.287	4.098	4.767	5.160	5.891
0.001	4.096	4.807	5.435	5.811	6.521

Table 4.5. δ, $B(\delta)$, and $MSE(\delta)$ for $n = 15$, $\alpha = .50$, and $(\sigma_y^2, \sigma_x^2) = (6,12)$

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a]	0.3577	0.7153	1.0730	1.4307	2.1460	2.8613
	0.0000[b]	-0.0134	-0.0225	-0.0252	-0.0223	-0.0104	-0.0027
	0.3780[c]	0.3843	0.3996	0.4157	0.4253	0.4210	0.4076
-1/3	0.0000	0.3981	0.7963	1.1944	1.5926	2.3889	3.1852
	0.0000	-0.0124	-0.0199	-0.0208	-0.0167	-0.0059	-0.0010
	0.3843	0.3898	0.4025	0.4141	0.4190	0.4115	0.4028
0	0.0000	0.4564	0.9129	1.3693	1.8257	2.7386	3.6515
	0.0000	-0.0108	-0.0163	-0.0152	-0.0104	-0.0023	-0.0002
	0.3905	0.3948	0.4039	0.4105	0.4112	0.4041	0.4005
1/3	0.0000	0.5512	1.1024	1.6535	2.2047	3.3071	4.4094
	0.0000	-0.0080	-0.0105	-0.0079	-0.0040	-0.0004	-0.0000
	0.3961	0.3986	0.4029	0.4047	0.4035	0.4005	0.4000
1/2	0.0000	0.6278	1.2556	1.8834	2.5112	3.7668	5.0224
	0.0000	-0.0055	-0.0064	-0.0039	-0.0015	-0.0001	0.0000
	0.3985	0.3997	0.4015	0.4018	0.4010	0.4001	0.4000
2/3	0.0000	0.7489	1.4979	2.2468	2.9957	4.4936	5.9914
	0.0000	-0.0014	-0.0013	-0.0005	-0.0001	0.0000	0.0000
	0.3999	0.4000	0.4001	0.4001	0.4000	0.4000	0.4000
3/4	0.0000	0.8434	1.6868	2.5301	3.3735	5.0603	6.7471
	0.0000	0.0018	0.0013	0.0004	0.0001	0.0000	0.0000
	0.3999	0.4000	0.4001	0.4001	0.4000	0.4000	0.4000
7/8	0.0000	1.0910	2.1819	3.2729	4.3638	6.5457	8.7277
	0.0000	0.0094	0.0037	0.0004	0.0000	0.0000	0.0000
	0.3969	0.4023	0.4028	0.4004	0.4000	0.4000	0.4000

[a] δ

[b] $B(\delta)$

[c] $MSE(\delta)$

Table 4.6. δ, B(δ), and MSE(δ) for $n = 15$, $\alpha = .25$, and $(\sigma_y^2, \sigma_x^2) = (6,12)$

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a]	0.3577	0.7153	1.0730	1.4307	2.1460	2.8613
	0.0000[b]	-0.0524	-0.0905	-0.1064	-0.1006	-0.0558	-0.0180
	0.3146	0.3369	0.3930	0.4569	0.5026	0.5044	0.4480
-1/3	0.0000	0.3981	0.7963	1.1944	1.5926	2.3889	3.1852
	0.0000	-0.0486	-0.0811	-0.0899	-0.0782	-0.0339	-0.0077
	0.3393	0.3588	0.4056	0.4536	0.4805	0.4614	0.4196
0	0.0000	0.4564	0.9129	1.3693	1.8257	2.7386	3.6515
	0.0000	-0.0426	-0.0671	-0.0676	-0.0513	-0.0149	-0.0019
	0.3632	0.3784	0.4126	0.4418	0.4505	0.4248	0.4043
1/3	0.0000	0.5512	1.1024	1.6535	2.2047	3.3071	4.4094
	0.0000	-0.0317	-0.0447	-0.0374	-0.0217	-0.0029	-0.0001
	0.3850	0.3938	0.4108	0.4203	0.4177	0.4038	0.4002
1/2	0.0000	0.6278	1.2556	1.8834	2.5113	3.7668	5.0224
	0.0000	-0.0220	-0.0280	-0.0196	-0.0088	-0.0005	0.0000
	0.3940	0.3984	0.4059	0.4081	0.4053	0.4005	0.4000
2/3	0.0000	0.7489	1.4979	2.2468	2.9957	4.4936	5.9914
	0.0000	-0.0057	-0.0060	-0.0030	-0.0008	0.0000	0.0000
	0.3997	0.4000	0.4004	0.4004	0.4001	0.4000	0.4000
3/4	0.0000	0.8434	1.6868	2.5301	3.3735	5.0603	6.7471
	0.0000	0.0073	0.0064	0.0024	0.0004	0.0000	0.0000
	0.3995	0.4001	0.4006	0.4004	0.4001	0.4000	0.4000
7/8	0.0000	1.0910	2.1819	3.2729	4.3638	6.5457	8.7277
	0.0000	0.0397	0.0199	0.0028	0.0001	0.0000	0.0000
	0.3881	0.4083	0.4142	0.4032	0.4002	0.4000	0.4000

[a] δ

[b] B(δ)

[c] MSE(δ)

Table 4.7. δ, $B(\delta)$, and $MSE(\delta)$ for $n = 15$, $\alpha = .10$, and $(\sigma_y^2, \sigma_x^2) = (6,12)$

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a]	0.3577	0.7153	1.0730	1.4307	2.1460	2.8613
	0.0000[b]	-0.1077	-0.1937	-0.2425	-0.2495	-0.1721	-0.0722
	0.2267[c]	0.2658	0.3694	0.5016	0.6180	0.6892	0.5771
-1/3	0.0000	0.3981	0.7963	1.1944	1.5926	2.3889	3.1852
	0.0000	-0.1002	-0.1756	-0.2103	-0.2026	-0.1140	-0.0351
	0.2768	0.3110	0.3989	0.5030	0.5817	0.5868	0.4828
0	0.0000	0.4564	0.9129	1.3693	0.8257	2.7386	3.6515
	0.0000	-0.0883	-0.1483	-0.1650	-0.1424	-0.0570	-0.0104
	0.3252	0.3522	0.4180	0.4867	0.5242	0.4868	0.4226
1/3	0.0000	0.5512	1.1024	1.6535	2.2047	3.3071	4.4094
	0.0000	-0.0663	-0.1026	-0.0986	-0.0684	-0.0138	-0.0009
	0.3695	0.3852	0.4197	0.4469	0.4500	0.4170	0.4015
1/2	0.0000	0.6278	1.2556	1.8834	2.5112	3.7668	5.0224
	0.0000	-0.0465	-0.0665	-0.0554	-0.0310	-0.0032	-0.0001
	0.3879	0.3958	0.4116	0.4204	0.4170	0.4029	0.4001
2/3	0.0000	0.7489	1.4979	2.2468	2.9957	4.4936	5.9914
	0.0000	-0.0123	-0.0152	-0.0097	-0.0036	-0.0001	0.0000
	0.3993	0.3999	0.4009	0.4011	0.4006	0.4000	0.4000
3/4	0.0000	0.8434	1.6868	2.5301	3.3735	5.0603	6.7471
	0.0000	0.0159	0.0171	0.0084	0.0021	0.0000	0.0000
	0.3991	0.4001	0.4015	0.4013	0.4005	0.4000	0.4000
7/8	0.0000	1.0910	2.1819	3.2729	4.3638	6.5457	8.7277
	0.0000	0.0909	0.0621	0.0131	0.0009	0.0000	0.0000
	0.3759	0.4150	0.4398	0.4142	0.4014	0.4000	0.4000

[a] δ

[b] $B(\delta)$

[c] $MSE(\delta)$

Table 4.8. δ, $B(\delta)$, and $MSE(\delta)$ for n=15, α = .50, and (σ_y^2, σ_x^2) = (6, 24)

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a] 0.0000[b] 0.3797[c]	0.2855 -0.0105 0.3835	0.5710 -0.0187 0.3934	0.8566 -0.0233 0.4058	1.1421 -0.0240 0.4167	1.7131 -0.0171 0.4245	2.2841 -0.0080 0.4168
-1/3	0.0000 0.0000 0.3875	0.3141 -0.0090 0.3903	0.6283 -0.0157 0.3974	0.9424 -0.0188 0.4058	1.2566 -0.0183 0.4123	1.8849 -0.0111 0.4143	2.5132 -0.0042 0.4077
0	0.0000 0.0000 0.3943	0.3536 -0.0067 0.3959	0.7071 -0.0114 0.3998	1.0607 -0.0128 0.4039	1.4142 -0.0115 0.4065	2.1213 -0.0055 0.4056	2.8284 -0.0015 0.4021
1/3	0.0000 0.0000 0.3991	0.4129 -0.0030 0.3995	0.8257 -0.0048 0.4002	1.2386 -0.0048 0.4008	1.6514 -0.0037 0.4011	2.4771 -0.0012 0.4006	3.3028 -0.0002 0.4001
1/2	0.0000 0.0000 0.4000	0.4564 0.0000 0.4000	0.9129 0.0000 0.4000	1.3693 0.0000 0.4000	1.8257 0.0000 0.4000	2.7386 0.0000 0.4000	3.6515 0.0000 0.4000
2/3	0.0000 0.0000 0.3986	0.5176 0.0045 0.3994	1.0351 0.0063 0.4009	1.5527 0.0051 0.4016	2.0703 0.0029 0.4014	3.1054 0.0004 0.4003	4.1405 0.0000 0.4000
3/4	0.0000 0.0000 0.3964	0.5590 0.0077 0.3988	1.1180 0.0101 0.4028	1.6771 0.0074 0.4043	2.2361 0.0036 0.4031	3.3541 0.0003 0.4004	4.4721 0.0000 0.4000
7/8	0.0000 0.0000 0.3893	0.6455 0.0147 0.3982	1.2910 0.0167 0.4109	0.9365 0.0097 0.4119	2.5820 0.0034 0.4060	3.8730 0.0001 0.4003	5.1640 0.0000 0.4000

[a] δ

[b] $B(\delta)$

[c] $MSE(\delta)$

Table 4.9. δ, B(δ), and MSE(δ) for n = 15, α = .25, and (σ_y^2, σ_x^2) = (6,24)

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a]	0.2855	0.5710	0.8566	1.1421	1.7131	2.2841
	0.0000[b]	-0.0408	-0.0743	-0.0955	-0.1024	-0.0820	-0.0448
	0.3215[c]	0.3348	0.3704	0.4174	0.4623	0.5074	0.4870
-1/3	0.0000	0.3141	0.6283	0.9424	1.2566	1.8849	2.5132
	0.0000	-0.0350	-0.0626	-0.0779	-0.0797	-0.0557	-0.0251
	0.3515	0.3614	0.3872	0.4196	0.4476	0.4659	0.4431
0	0.0000	0.3536	0.7071	1.0607	1.4142	2.1213	2.8284
	0.0000	-0.0264	-0.0457	-0.0541	-0.0515	-0.0293	-0.0097
	0.3779	0.3835	0.3978	0.4142	0.4262	0.4274	0.4131
1/3	0.0000	0.4129	0.8257	1.2386	1.6514	2.4771	3.3028
	0.0000	-0.0118	-0.0194	-0.0210	-0.0177	-0.0070	-0.0014
	0.3966	0.3978	0.4005	0.4032	0.4045	0.4031	0.4009
1/2	0.0000	0.4564	0.9129	1.3693	1.8257	2.7386	3.6515
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000
2/3	0.0000	0.5176	1.0351	1.5527	2.0703	3.1054	4.1405
	0.0000	0.0178	0.0263	0.0236	0.0152	0.0027	0.0002
	0.3947	0.3975	0.4031	0.4069	0.4067	0.4020	0.4002
3/4	0.0000	0.5590	1.1180	1.6771	2.2361	3.3541	4.4721
	0.0000	0.0307	0.0430	0.0353	0.0200	0.0025	0.0001
	0.3862	0.3945	0.4104	0.4188	0.4159	0.4032	0.4002
7/8	0.0000	0.6455	1.2910	1.9365	2.5820	3.8730	5.1640
	0.0000	0.0590	0.0732	0.0490	0.0207	0.0011	0.0000
	0.3585	0.3905	0.4428	0.4557	0.4341	0.4028	0.4000

[a] δ

[b] B(δ)

[c] MSE(δ)

Table 4.10. δ, B(δ), and MSE(δ) for n = 15, α = .10, and $(\sigma_y^2, \sigma_x^2) = (6,24)$

ρ	Δ=0	Δ=0.5	Δ=1.0	Δ=1.5	Δ=2.0	Δ=3.0	Δ=4.0
-2/3	0.0000[a]	0.2855	0.5710	0.8566	1.1421	1.7131	2.2841
	0.0000[b]	-0.0835	-0.1561	-0.2089	-0.2369	-0.2200	-0.1449
	0.2407[c]	0.2639	0.3281	0.4190	0.5164	0.6533	0.6542
-1/3	0.0000	0.3141	0.6283	0.9424	1.2566	1.8849	2.5132
	0.0000	-0.0718	-0.1324	-0.1730	-0.1893	-0.1577	-0.0882
	0.3016	0.3189	0.3659	0.4300	0.4942	0.5655	0.5380
0	0.0000	0.3536	0.7071	1.0607	1.4142	2.1213	2.8284
	0.0000	-0.0542	-0.0977	-0.1229	-0.1272	-0.0895	-0.0386
	0.3551	0.3650	0.3913	0.4251	0.4553	0.4752	0.4475
1/3	0.0000	0.4129	0.8257	1.2386	1.6514	2.4771	3.3028
	0.0000	-0.0243	-0.0422	-0.0497	-0.0466	-0.0242	-0.0066
	0.3932	0.3952	0.4004	0.4063	0.4104	0.4098	0.4038
1/2	0.0000	0.4564	0.9129	1.3693	1.8257	2.7386	3.6515
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000
2/3	0.0000	0.5176	1.0351	1.5527	2.0703	3.1054	4.1405
	0.0000	0.0372	0.0594	0.0604	0.0455	0.0119	0.0011
	0.3893	0.3942	0.4054	0.4153	0.4181	0.4081	0.4011
3/4	0.0000	0.5590	1.1180	1.6771	2.2361	3.3541	4.4721
	0.0000	0.0643	0.0988	0.0937	0.0637	0.0121	0.0007
	0.3720	0.3868	0.4191	0.4437	0.4456	0.4145	0.4012
7/8	0.0000	0.6455	1.2910	1.9365	2.5820	3.8730	5.1640
	0.0000	0.1250	0.1753	0.1410	0.0748	0.0065	0.0001
	0.3159	0.3736	0.4859	0.5425	0.5121	0.4160	0.4004

[a] δ

[b] B(δ)

[c] MSE(δ)

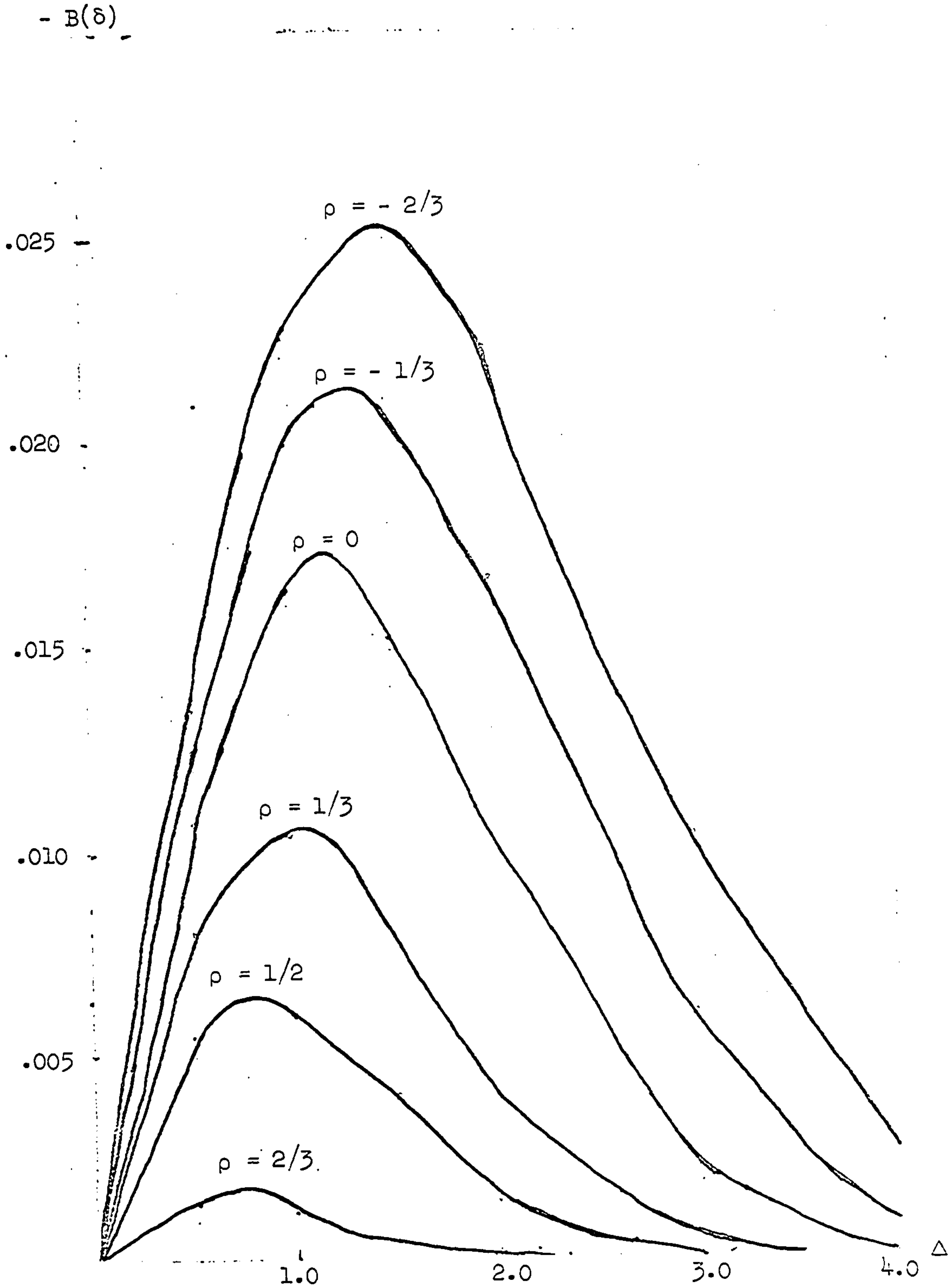

Figure 4.5. $-B(\delta)$ for $(\sigma_y^2, \sigma_x^2) = (6, 12)$, $\alpha = .50$, $n = 15$

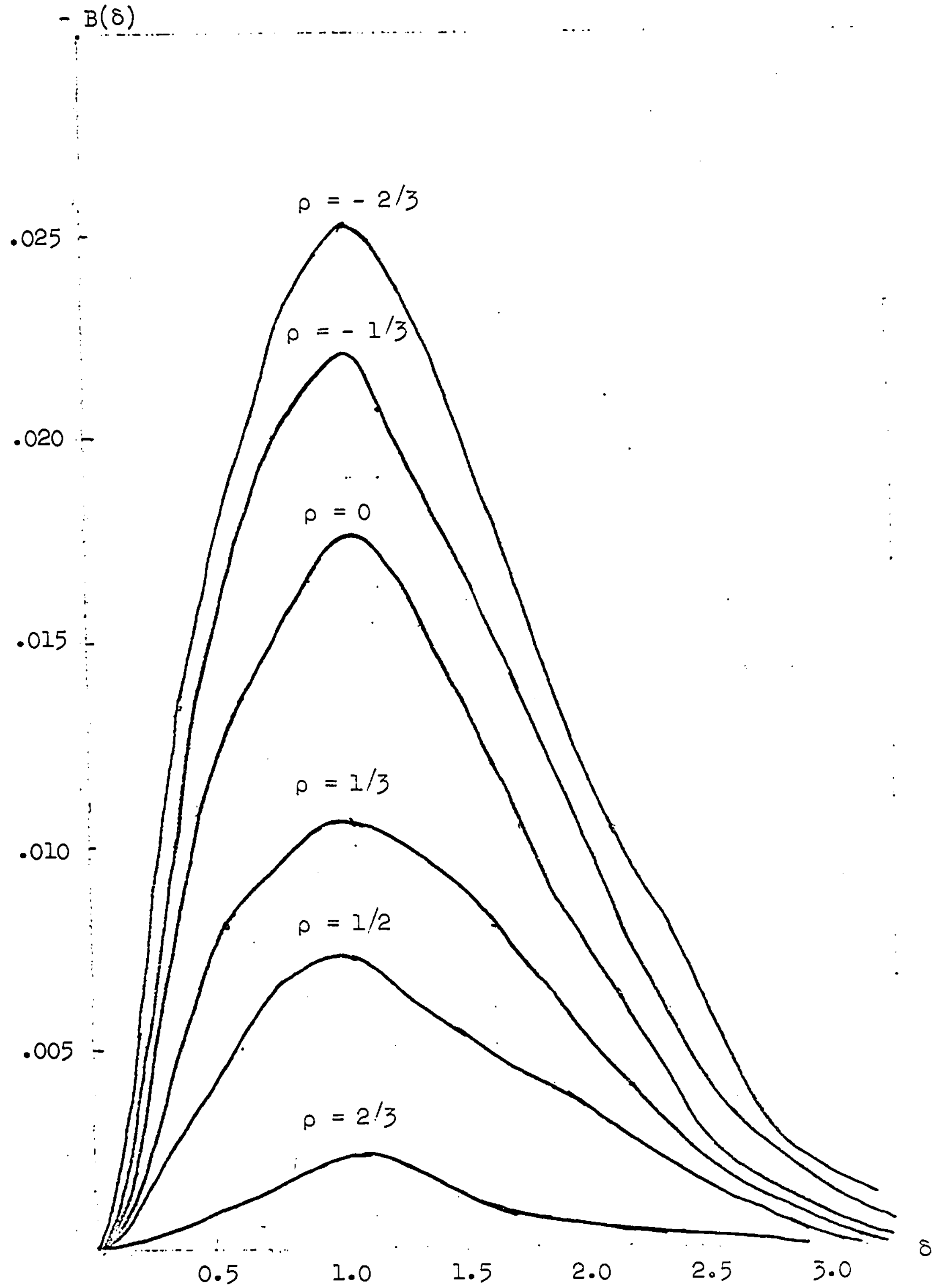

Figure 4.6. $-B(\delta)$ for $(\sigma_y^2, \sigma_x^2) = (6, 12)$, $\alpha = .50$, $n = 15$

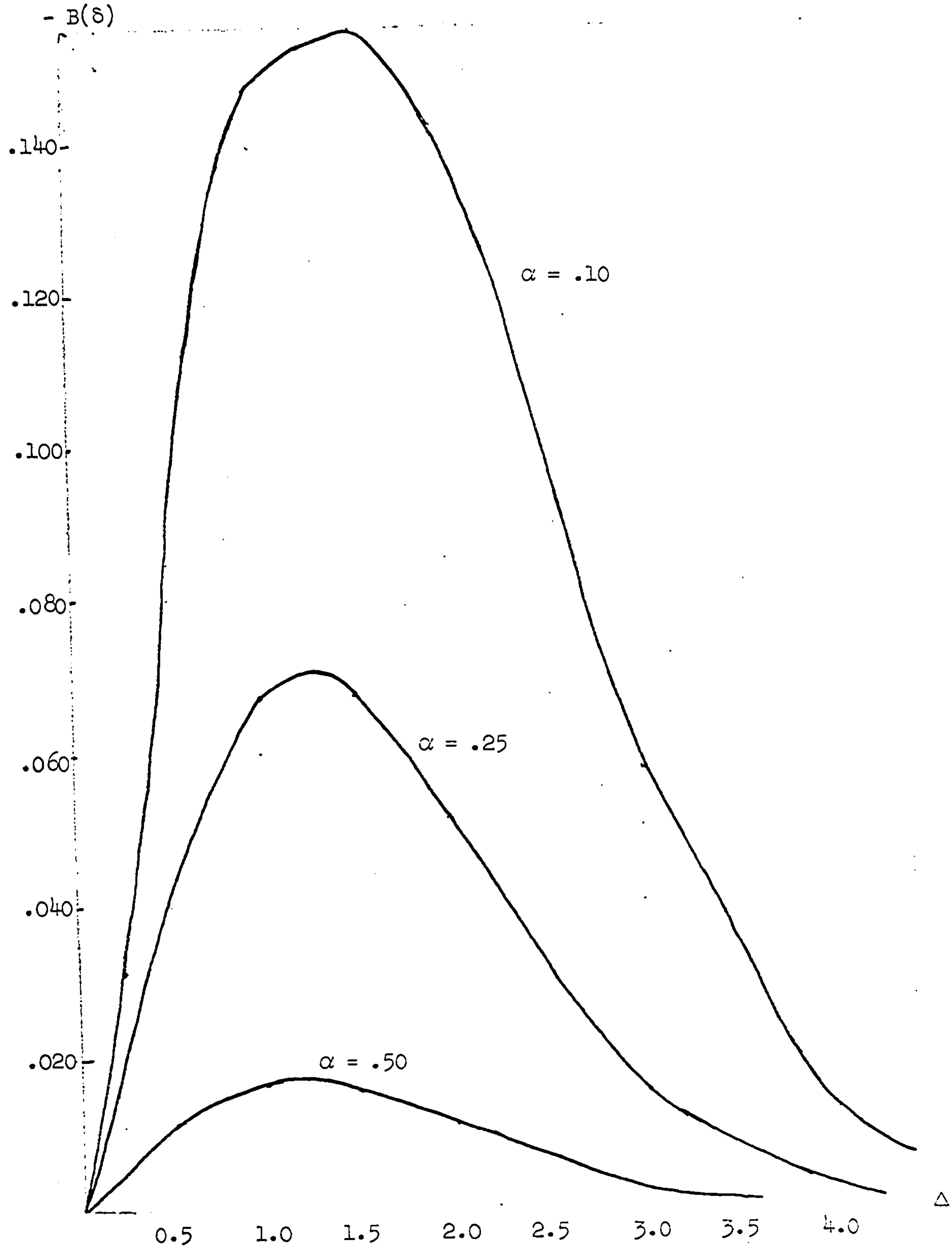

Figure 4.7. $-B(\delta)$ for $(\sigma_y^2, \sigma_x^2) = (6, 12)$, $n = 15$, $\rho = 0$

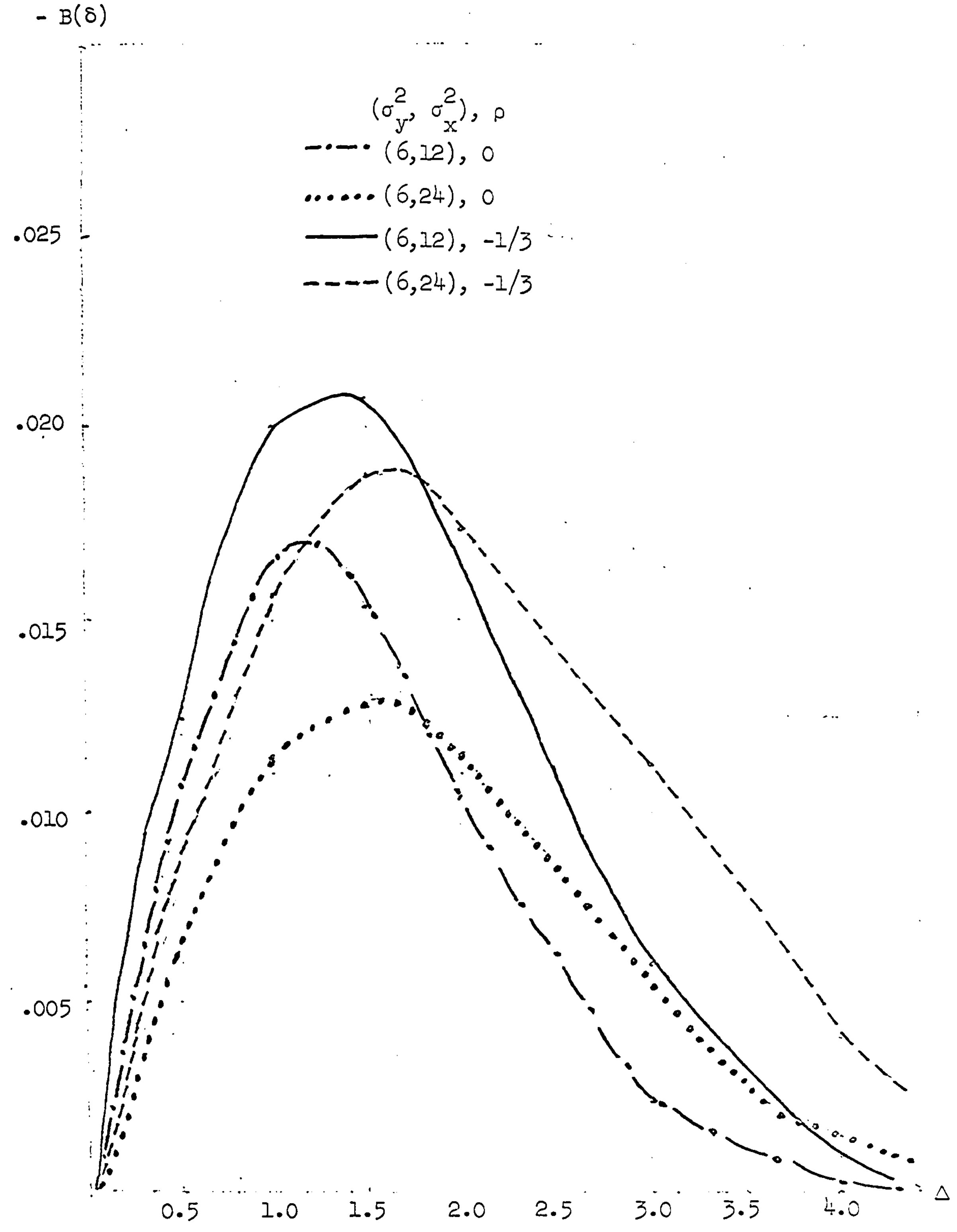

Figure 4.8. $-B(\delta)$ for $n = 15$ and $\alpha = .50$

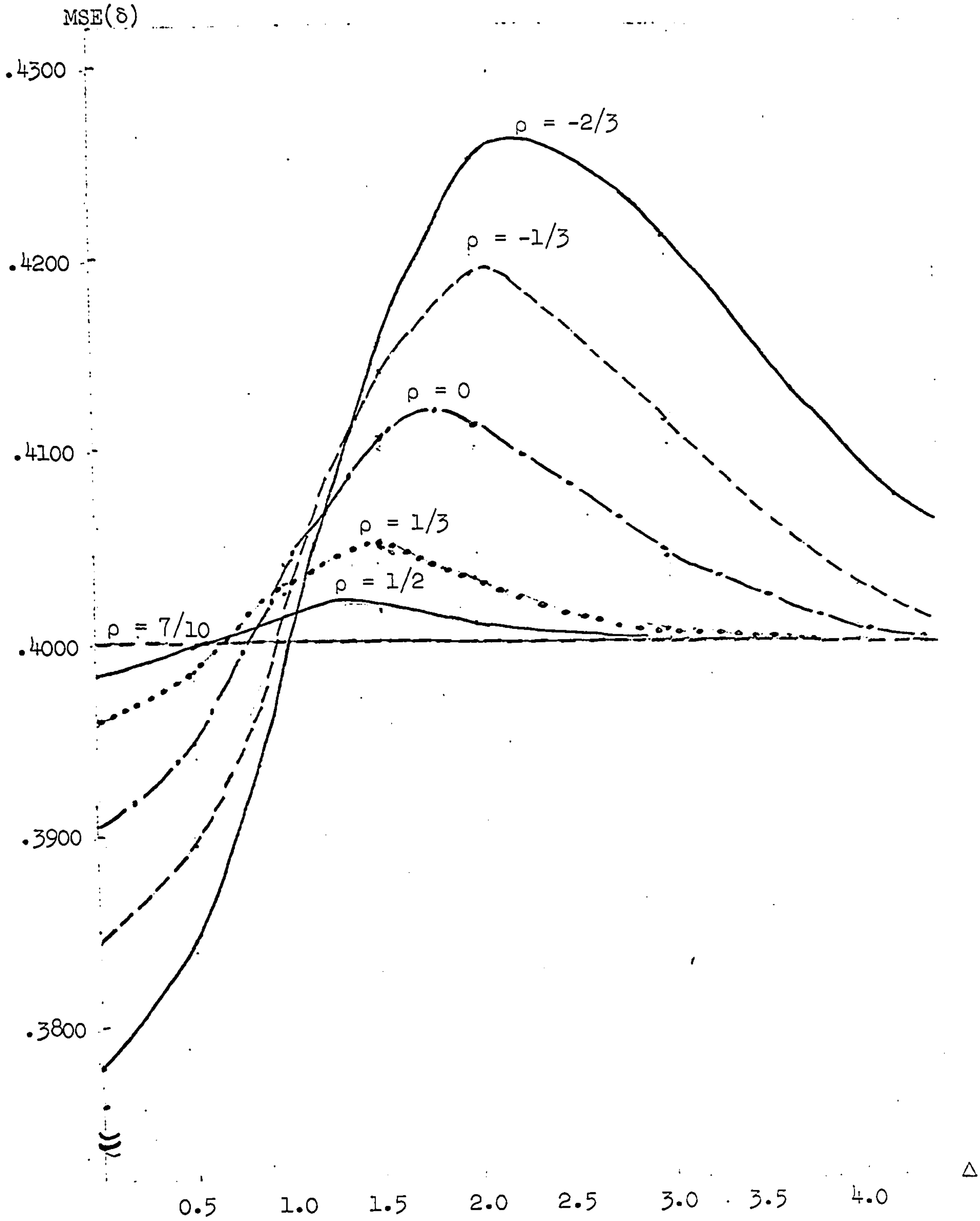

Figure 4.9. MSE(δ) for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\alpha = .50$, $n = 15$

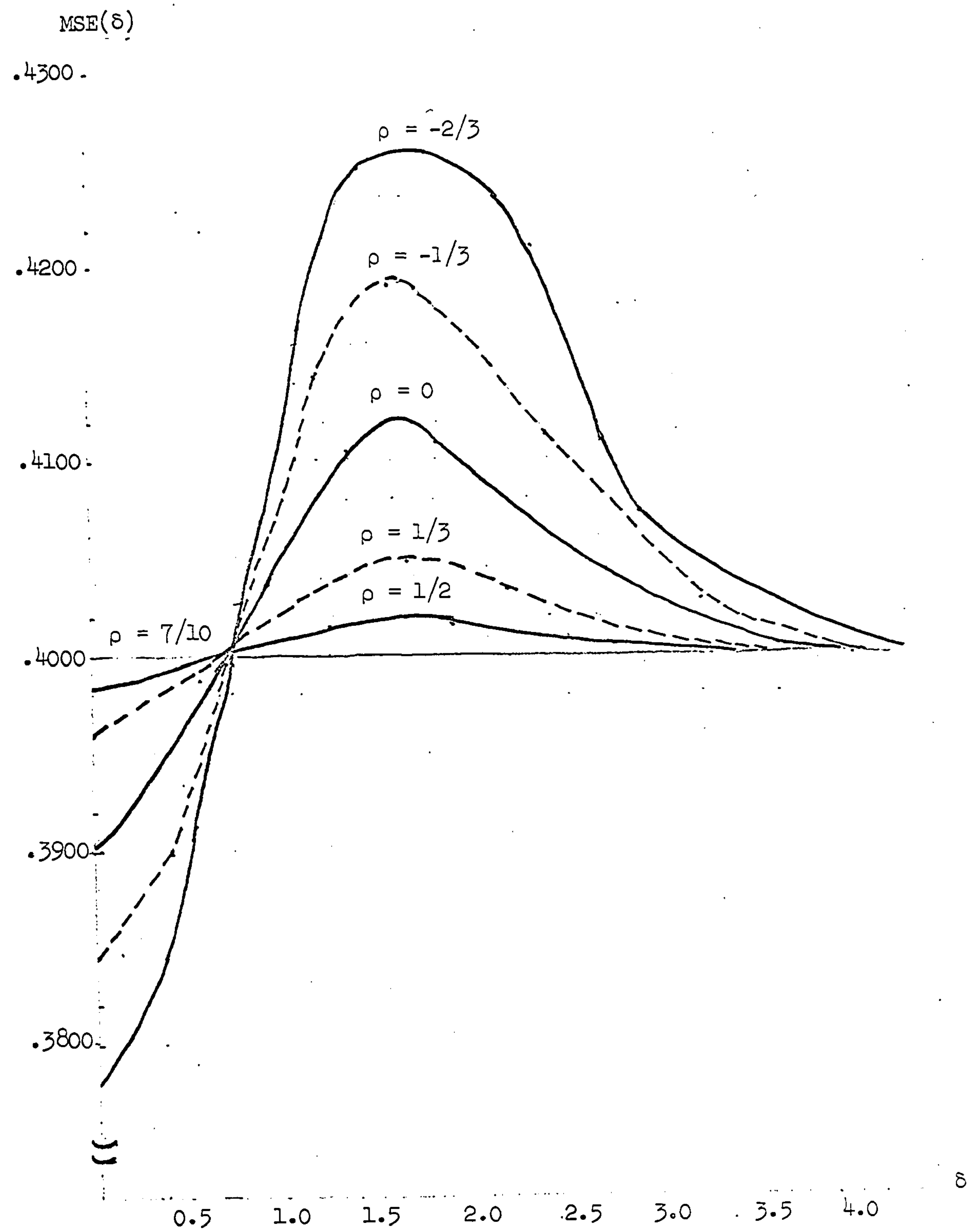

Figure 4.10. MSE(δ) for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\alpha = .50$, $n = 15$

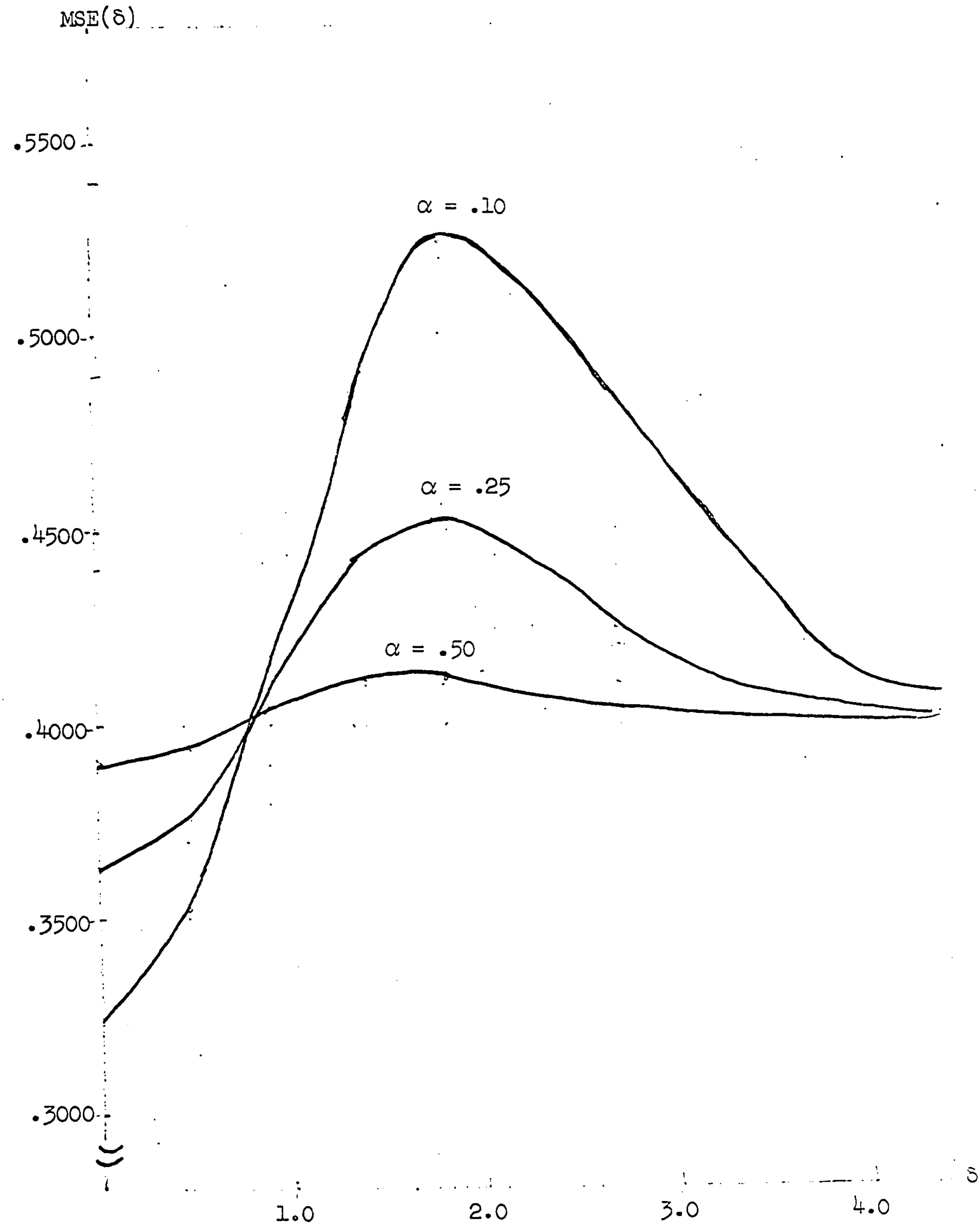

Figure 4.11. MSE(δ) for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $n = 15$, $\rho = 0$

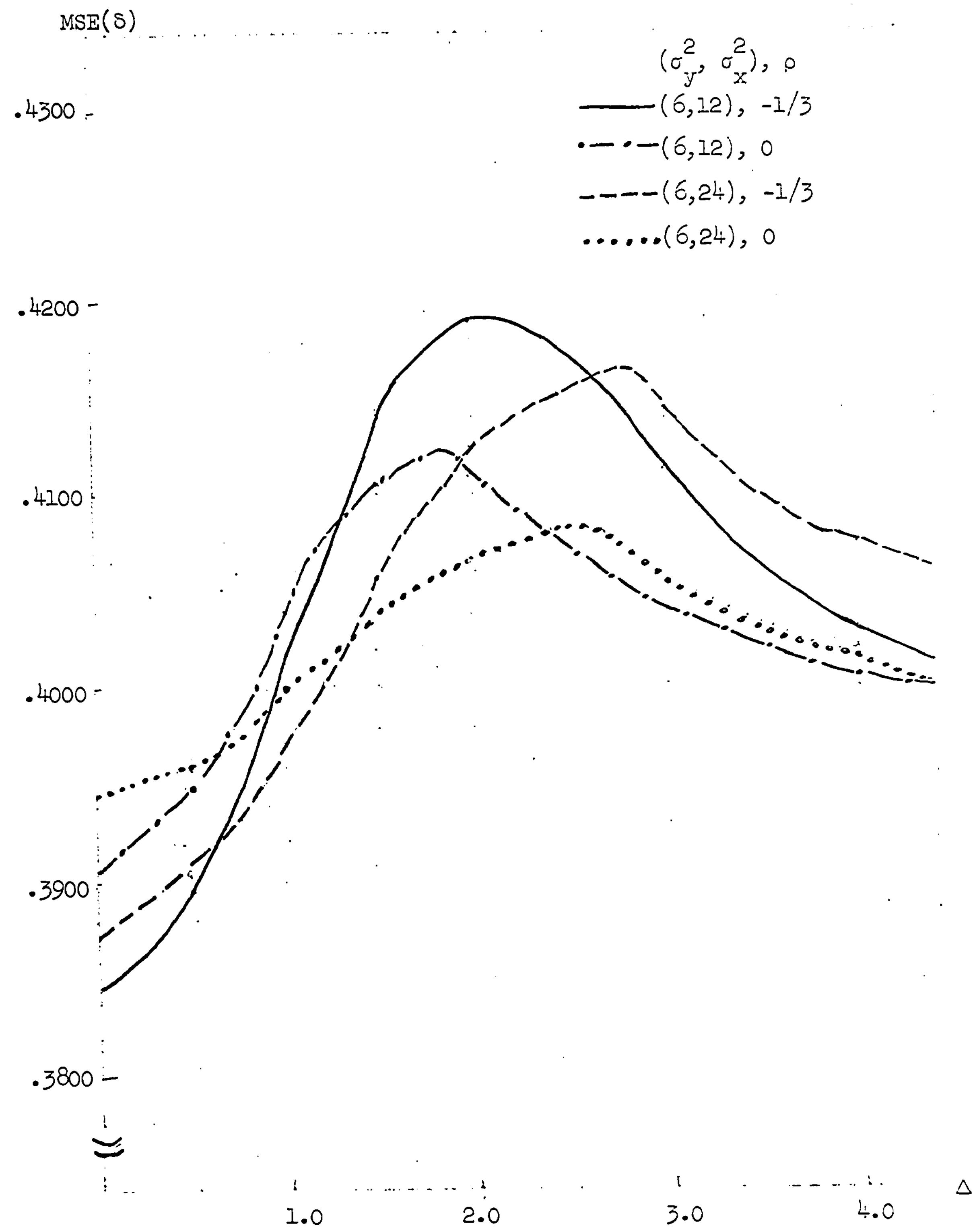

Figure 4.12. MSE(δ) for n = 15 and α = .50

V. SOME NUMERICAL INVESTIGATIONS ON ONE AND TWO STAGE PRELIMINARY TEST PROCEDURES

A. A Comparison of Three Different Weights for the Preliminary Test Procedure

1. The optimum weights

In section D of Chapter IV the optimum values for the weights w_1 and w_2 were derived for the special case when a one stage bivariate sample of size n is drawn from a bivariate normal population with parameters (μ_y, μ_x, ρ, σ_y^2, σ_x^2) and the preliminary test procedure is used to estimate μ_y. These optimum weights, denoted by w_{20} and w_{10}, are

$$w_{20} = \frac{k_1 A(\delta)}{(k_1+k_2)C(\delta)} \tag{5.1.1}$$

$$w_{10} = 1 - w_{20},$$

where

$$\begin{aligned} k_2 &= (\sigma_x^2 - \rho\sigma_x\sigma_y)/n \\ k_1 &= (\sigma_y^2 - \rho\sigma_x\sigma_y)/n \\ \sigma_z^2 &= k_1 + k_2 \\ \delta &= \Delta/\sigma_z \end{aligned} \tag{5.1.2}$$

and

$$A(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} t(t+\delta)\, \phi(t)\, dt \tag{5.1.3}$$

$$C(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (t+\delta)^2\, \phi(t)\, dt. \tag{5.1.4}$$

$\phi(t)$ is the $N(0, 1)$ density, and Δ and ξ_α are defined in equations (3.7.1) and (3.7.3). Alternative expressions for $A(\delta)$ and $C(\delta)$, appropriate for hand calculation with a book of standard tables, are

$$A(\delta) = \left[\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right] - \xi_\alpha\left[\phi(\xi_\alpha-\delta) + \phi(-\xi_\alpha-\delta)\right] \tag{5.1.5}$$

and

$$C(\delta) = (\delta^2+1)\left[\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right] - \xi_\alpha\left[\phi(\xi_\alpha-\delta) + \phi(-\xi_\alpha-\delta)\right] - \delta\left[\phi(\xi_\alpha-\delta) - \phi(-\xi_\alpha-\delta)\right], \tag{5.1.6}$$

where $\Phi(t)$ is the cumulative distribution function of the $N(0, 1)$ distribution. Recall that $MSE(\delta)$, with $w_2 = w_{20}$, is always less than σ_y^2/n for any value of Δ or δ.

However, δ is unknown, and thus w_{20} cannot be used directly. Section D of Chapter IV discusses several possibilities for approximating w_{20}. One of these possibilities, the estimation of δ from the sample, is discussed in this section. Recall from Chapter IV that the covariance matrix elements ρ, σ_y^2, and σ_x^2 are assumed to be known.

2. The estimator J

An unbiased estimate of δ from the sample is $\hat{\delta}$, where

$$\hat{\delta} = (\bar{y}_n - \bar{x}_n)/\sigma_z. \tag{5.1.7}$$

Let the estimator J be obtained by following the preliminary test procedure described in Chapter IV, where now the weights w_1 and w_2 are obtained by replacing δ by $\hat{\delta}$ in equation (5.1.1). Thus,

$$w_{2(J)} = \frac{k_1 A(\hat{\delta})}{(k_1+k_2)C(\hat{\delta})}$$

$$w_{1(J)} = 1 - w_{2(J)}, \tag{5.1.8}$$

where the functions $A(\delta)$ and $C(\delta)$ are defined in equations (5.1.3) and (5.1.4). If H_o: $\mu_y = \mu_x$ is accepted by the preliminary test, then the pooled estimator $w_{1(J)}\bar{y}_n + w_{2(J)}\bar{x}_n$ is used. If H_A: $\mu_y \neq \mu_x$ is accepted by the preliminary test, then the non-pooled estimator $\bar{y}_n$ is used. The motivation for suggesting the estimator J is that if $\hat{\delta}$ is close to the population value δ, then MSE(J) should be less than σ_y^2/n, whatever the value of δ in the population. This, then, would be an improvement over the preliminary test scheme discussed in Chapter IV which can have a mean square error considerably greater than σ_y^2/n for some values of δ.

The derivation of the bias B(J) and mean square error MSE(J) is very complicated since, in the pooled estimator $w_{1(J)}\bar{y}_n + w_{2(J)}\bar{x}_n$, the random variable $\hat{\delta}$ appears as the argument of the cumulative normal distribution function. No attempt was made to evaluate B(J) and MSE(J) analytically.

Results of a Monte Carlo study on the estimator J are reported at the end of this section.

3. <u>The estimator K</u>

Although the estimator J can be computed with the aid of standard tables, two approximations to J, the estimators K and L, were also considered. Note that $A(\delta)$ and $C(\delta)$ contain the terms $\phi(t)$ and $\Phi(t)$ evaluated either at $t = (\xi_\alpha - \delta)$ or at $t = (-\xi_\alpha - \delta)$. Now expand $\phi(t)$ and $\Phi(t)$ in a Taylor series around the point $-\delta$, where $-\delta$ is the midpoint of the interval $(-\xi_\alpha - \delta,\ \xi_\alpha - \delta)$. This yields

$$\phi(x) \doteq \phi(-\delta)\left[1 + \delta(x+\delta) + \frac{(\delta^2-1)(x+\delta)^2}{2!} + \frac{(\delta^3-3\delta)(x+\delta)^3}{3!} + \frac{(\delta^4-6\delta^2+3)(x+\delta)^4}{4!}\right] \qquad (5.1.9)$$

and

$$\Phi(x) \doteq \Phi(-\delta) + \phi(-\delta)\left[(x+\delta) + \frac{\delta(x+\delta)^2}{2!} + \frac{(\delta^2-1)(x+\delta)^3}{3!} + \frac{(\delta^3-3\delta)(x+\delta)^4}{4!}\right]. \qquad (5.1.10)$$

Equations (5.1.9) and (5.1.10) are then used to evaluate the terms $\phi(\xi_\alpha - \delta)$, $\phi(-\xi_\alpha - \delta)$, $\Phi(-\xi_\alpha - \delta)$, and $\Phi(\xi_\alpha - \delta)$ which occur in $A(\delta)$ and $C(\delta)$. Retaining the first five terms of the Taylor series expansion for the preceding four quantities, $A(\delta)/C(\delta)$ is approximated by

$$\frac{A(\delta)}{C(\delta)} \doteq T(\delta) = \frac{\left[\frac{(1-\delta^2)}{3} - \frac{\xi_\alpha^2(\delta^4-6\delta^2+3)}{4!}\right]}{\left[\frac{1}{3} - \frac{\xi_\alpha^2(\delta^4-6\delta^2+3)}{4!}\right]}. \tag{5.1.11}$$

Let the estimator K be obtained by following the preliminary test procedure in Chapter IV where now the weights $w_{1(K)}$ and $w_{2(K)}$ for the pooled estimator are

$$w_{2(K)} = k_1 T(\hat{\delta})/(K_1+k_2) \tag{5.1.12}$$

$$w_{1(K)} = 1 - w_{2(K)}.$$

B(K) and MSE(K) are analytically intractable, and the results of a Monte Carlo study on the estimator K are reported at the end of this section.

4. The estimator L

The estimator L is obtained by following the preliminary test procedure in Chapter IV where the weights for the pooled estimator are

$$\begin{aligned} w_{2(L)} &= k_1 Y(\hat{\delta})/(k_1+k_2) \\ w_{1(L)} &= 1 - w_{2(L)} \end{aligned} \tag{5.1.13}$$

and

$$A(\delta)/C(\delta) \doteq Y(\delta) = 1 - \delta^2. \tag{5.1.14}$$

$Y(\delta)$ is obtained as a further approximation to $A(\delta)/C(\delta)$ by retaining less terms in the Taylor series expansions of $\phi(\xi_\alpha-\delta)$, $\phi(-\xi_\alpha-\delta)$, $\Phi(\xi_\alpha-\delta)$, and $\phi(-\xi_\alpha-\delta)$. The bias and mean square error of the estimator L were investigated by Monte Carlo methods, and the results are presented in the next section.

5. Monte Carlo comparisons

The bivariate normal population with parameters $(\sigma_y^2, \sigma_x^2) = (6, 12)$ and $\rho = 1/3$ was chosen for the Monte Carlo comparison of the estimators J, K, and L. Sample sizes $n = 4, 9, 15$ and α-levels .10, .25, and .50 were considered for various values of Δ ranging from 0 to 4.0.

The bivariate sample means $(\bar{x}_n, \bar{y}_n)$ were generated directly by generating bivariate normal random variables $(\tilde{z}_x, \tilde{z}_y)$ such that

$$\begin{aligned}
&E(\tilde{z}_x) = -\Delta \\
&E(\tilde{z}_y) = 0 \\
&V(\tilde{z}_x) = \sigma_x^2/n = 6/n \\
&V(\tilde{z}_y) = \sigma_y^2/n = 12/n \\
&\frac{\operatorname{cov}(\tilde{z}_x, \tilde{z}_y)}{\sqrt{V(\tilde{z}_x)V(\tilde{z}_y)}} = \rho = 1/3
\end{aligned} \tag{5.1.15}$$

The following three step procedure was repeated one thousand times: (1) a random bivariate normal sample mean $(\bar{x}_n, \bar{y}_n)$ was generated, (2) the preliminary test of $H_o: \mu_y = \mu_x$ versus $H_A: \mu_y \neq \mu_x$ was made using this bivariate sample mean, (3) the three estimators J, K, and L were each separately computed as described previously, using the sample mean $(\bar{x}_n, \bar{y}_n)$. Then, the mean, variance, and mean square error (variance plus square of bias) of each of J, K, and L were calculated using the one thousand different estimates which were produced. The means of J, K, and L give estimates of the respective biases since $\mu_y = 0$. The sample mean square errors give estimates of the true mean square errors of J, K, and L.

For $n = 15$ and $\alpha = .50, .25, .10$ the Monte Carlo bias and mean square error for J, K, and L were compared to those values in Tables 4.5 through 4.7 for $\rho = 1/3$ and various values of Δ. The Monte Carlo results are not tabled, but are merely summarized. For $\alpha = .10$, the biases of J, K, and L were all smaller than $B(\delta)$ in Table 4.7 for values of Δ ranging from .50 to 3.0, or equivalently δ ranging from .55 to 3.3. The biggest savings in bias occurs from $\delta = 1$ to $\delta = 2$. For example, for $\Delta = 1$, and hence $\delta = 1.1$, $B(\delta) = -0.1026$ from Table 4.7, and the Monte Carlo biases of the estimators J, K, and L were -0.001, -0.014, and 0.048, respectively. For intermediate values of δ, say around 1.5 to 2.5, the mean square errors of J, K, and L were all lower than $MSE(\delta)$. For example, for $\Delta = 2.0$, and hence $\delta = 2.2$, $MSE(J) = .412$, $MSE(K) = .419$, and $MSE(L) = .408$, while $MSE(\delta) = .450$ from Table 4.7. For larger values of δ around 3 and 4, the estimators J, K, and L behaved similarly to the preliminary test procedure. This is to be expected since, for large δ, all four procedures will mostly reject $H_o:\mu_y = \mu_x$ and hence all will use the estimator $\bar{y}_n$. For Δ around zero, the estimators J, K, and L all had considerable bias and mean square error compared to the regular preliminary test procedure. Of course, at

$\Delta = 0$ the optimum weights are exactly those used by the preliminary test procedure, and the estimators J, K, and L lose precision by using estimated weights. For example, at $\Delta = 0$, MSE(J)=.407, MSE(K)=.397, and MSE(L)=.454, whereas from Table 4.7, MSE(δ)=.370. For the comparison within the estimators J, K, and L, K seemed to have smallest mean square error for δ from 0 to 1, J seemed to have smallest mean square error for δ from 1 to 3, and for larger δ all three were equivalent.

As α increased to .25 and .50, there was less discrepancy in mean square error within the estimators J, K, and L and between them and the preliminary test procedure. For δ ranging from 1 to 2, the bias B(δ) was halved or quartered by J, K, and L. However, for high α, B(δ) is usually small anyway, and hence J, K, and L have no real advantage.

For n=4 and n=9, the same general results apply. Although B(δ) is generally higher for smaller n, the estimators J, K, and L do not do significantly better because they, in turn, are estimating δ with a small sample.

In summary then, the estimators J, K, and L have larger bias and mean square error than the preliminary test procedure for Δ near zero. They are better than the preliminary test procedure for moderate δ from 1 to 2 or 3, with a significant improvement in bias. For δ larger than 3, all four procedures behave similarly.

B. One Stage Preliminary Test Procedure Versus Two Stage Preliminary Test Procedure for Independent Random Variables

1. Objective of study

In Chapter III a two stage preliminary test estimation scheme was presented to estimate the mean μ_y of a bivariate normal population, and the bias $B(\delta)$ and mean square error $MSE(\delta)$ were derived for this scheme. The one stage scheme, of course, can be derived as a special case of the two stage scheme by letting $m_x = m_y = 0$, where m_x and m_y are the sizes of the second stage samples on $\tilde{x}$ and $\tilde{y}$, respectively.

The two stage plan requires sampling on two occasions since it is not known until after the preliminary test is done whether x or y variables will be sampled at the second stage. It seems plausible that the mean square errors for two stage plans should be less than the mean square errors for comparable one stage plans since the two stage plan appears to make more efficient use of the data. Since the formulas for bias and mean square error are complicated for both procedures, a direct comparison between the two approaches is not feasible. Thus, a numerical investigation was done to determine what advantages, if any, the two stage sampling plan has.

2. $B(\delta)$ and $MSE(\delta)$

For the numerical investigation discussed in this chapter, $\tilde{x}$ and $\tilde{y}$ are assumed to be independent, with σ_y^2 and σ_x^2 known. Thus, let ρ and n, the size of the bivariate sample in Chapter III, equal zero. Then the bias of the two stage preliminary test estimation procedure is easily obtained from equation (3.7.23) as

$$B(\delta) = \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \phi(t) \left[-w_2\delta\sigma_z + \frac{t}{\sigma_z}\left\{ \frac{-w_2\sigma_y^2}{n_y} - \frac{w_2\sigma_x^2}{N_x} + \frac{m_y\sigma_y^2}{n_yN_y} \right\} \right] dt \quad (5.2.1)$$

where, since n = 0,

$$\begin{aligned} k_1^* &= \sigma_y^2/n_y \\ k_2^* &= \sigma_x^2/n_x \\ \sigma_y^2 &= k_1^* + k_2^* \\ N_y &= n_y + m_y \\ N_x &= n_x + m_x, \end{aligned} \quad (5.2.2)$$

and n_x and n_y are the sizes of the first stage sample on $\tilde{x}$ and $\tilde{y}$, respectively.

The mean square error of the two stage procedure is easily obtained from equation (3.8.16) as

$$MSE(\delta) = \sigma_y^2/N_y + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[w_2^2\delta^2\sigma_z^2 + t^2A_2^* + 2\delta tA_1^* + A_0^* \right] \phi(t)\, dt \quad (5.2.3)$$

where, from equations (3.8.17) through (3.8.19) with n = 0,

$$A_2^* = \sigma_z^{-2}\left[\left\{ \frac{w_1\sigma_y^2}{n_y} - \frac{w_2\sigma_x^2}{N_x} \right\}^2 - \frac{\sigma_y^4}{N_y^2} \right] \quad (5.2.4)$$

$$A_1^* = w_2 \left\{ \frac{w_2 \sigma_x^2}{N_x} - \frac{w_1 \sigma_y^2}{n_y} \right\} \tag{5.2.5}$$

and

$$A_0^* = \left[\left\{ w_1 + \frac{w_2 n_x}{N_x} \right\}^2 - \frac{n_y^2}{N_y^2} \right] \frac{\sigma_y^2 \sigma_x^2}{n_y n_x \sigma_z^2} + \frac{w_2^2 m_x \sigma_x^2}{N_x^2} - \frac{m_y \sigma_y^2}{N_y^2} \tag{5.2.6}$$

Now, let the weights w_1 and w_2 in this section be chosen so as to minimize the variance of the pooled estimator $(w_1 \bar{y}_{n_y} + w_2 \bar{x}_{N_x})$. Thus,

$$\begin{aligned} w_2 &= \left[\sigma_y^2 / n_y \right] \Big/ \left[\sigma_x^2 / N_x + \sigma_y^2 / n_y \right] \\ w_1 &= 1 - w_2 . \end{aligned} \tag{5.2.7}$$

With the weights as in equation (5.2.7), then the bias is easily obtained from equation (5.2.1) as

$$B(\delta) = -\sigma_z \sigma_y^2 \int_{-\xi_\alpha - \delta}^{\xi_\alpha - \alpha} \left\{ \frac{\delta}{n_y \left[\sigma_x^2 / N_x + \sigma_y^2 / n_y \right]} + \frac{t}{N_y \sigma_z^2} \right\} \phi(t)\, dt. \tag{5.2.8}$$

Likewise, noting that A_1^* of equation (5.2.5) is zero with the weights in equation (5.2.7), MSE(δ) is readily obtained as

$$\mathrm{MSE}(\delta) = \sigma_y^2 / N_y + \int_{-\xi_\alpha - \delta}^{\xi_\alpha - \delta} (t^2 B_1 + B_0)\, \phi(t)\, dt \tag{5.2.9}$$

where

$$B_1 = -\sigma_y^4/\left[N_y^2\sigma_z^2\right] \tag{5.2.10}$$

and

$$B_0 = \frac{\sigma_y^4\delta^2\sigma_z^2}{n_y^2\left\{\dfrac{\sigma_x^2}{N_x} + \dfrac{\sigma_y^2}{n_y}\right\}^2} + \frac{\sigma_y^4 m_x \sigma_x^2}{n_y^2 N_x^2\left\{\dfrac{\sigma_x^2}{N_x} + \dfrac{\sigma_y^2}{n_y}\right\}^2} - \frac{m_y\sigma_y^2}{N_y^2}$$

$$+ \frac{\sigma_y^2\sigma_x^2}{n_y n_x \sigma_z^2}\ \frac{\left\{\dfrac{\sigma_x^2}{N_x} + \dfrac{n_x\sigma_y^2}{n_y N_x}\right\}^2}{\left\{\dfrac{\sigma_x^2}{N_x} + \dfrac{\sigma_y^2}{n_y}\right\}^2} - \frac{n_y^2}{N_y^2}\,. \tag{5.2.11}$$

In the numerical investigation discussed in the next section, equations (5.2.8) and (5.2.9) for $B(\delta)$ and $MSE(\delta)$ are used for all the calculations.

3. Numerical investigation

For the purpose of this investigation, a bivariate normal population with parameters $(\sigma_y^2, \sigma_x^2) = (6, 12)$ and $\rho = 0$ was chosen. Sampling costs $(c_y, c_x) = (2, 1)$ and $(c_y, c_x) = (4, 1)$ were considered as two representative cases. In both cases, a total sampling budget $C_0 = 36$ was allotted for sampling costs. The budget was allocated for the first and second stage samples as follows: (1) spend entire budget on first stage, (2) spend 1/3 of total budget on first stage, (3) spend 1/2 of total budget on first stage, and (4) spend 2/3 of total budget on first stage. Within each of these four cases, various selections of n_y and n_x were considered which used the allotted budget for the first stage. Of course, the second stage sample always samples either $\tilde{x}$ or $\tilde{y}$ with the remaining budget, and hence all second stage sample sizes m_x or m_y will be the same within each of the four main cases cited above.

Preliminary calculations as described above indicate that the two stage procedure has larger bias and mean square error than the one stage procedure. The bias and mean square error of the two stage procedure decrease as the amount of money spent on the first stage sample increases. This behavior may be caused by the lack of power of the preliminary test. σ_z^2 is generally smaller in the two stage procedure than in the one stage procedure because less sample observations are available. Hence, the preliminary test on the two stage procedure will accept $H_0:\mu_y = \mu_x$ more often, after which more x variables are sampled and included in the estimator. Hence, it seems to have more opportunity to be biased than the one stage preliminary test estimator.

At $\Delta = 0$, however, the bias of both the one and two stage procedures is zero, and in these cases the mean square error for the two stage procedure was less than the mean square error for the one stage procedure. Also, the mean square error of the two stage procedure for $\Delta = 0$ tended to decrease as the size of the first stage sample decreased.

A larger budget of $C_0 = 120$ was also allocated in a similar manner, and the same results were obtained as above for $C_0 = 36$.

VI. A PRELIMINARY TEST ESTIMATION SCHEME WITH REGRESSION ESTIMATOR

A. Statement of Problem

In the estimation scheme discussed in Chapter III, the non-pooled estimator $\hat{\mu}_y = \bar{y}_{n+n_y+m_y}$ was used whenever $H_A: \mu_y \neq \mu_x$ was accepted. This estimator utilizes none of the existing information on μ_x. In an effort to incorporate this extra information, consider now a one stage sampling procedure where a regression estimator is used whenever the pooled estimator cannot be used.

B. The Sampling Scheme

Let $(\tilde{x},\tilde{y})$ have a bivariate distribution with mean vector

$$E\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} = \begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix} \tag{6.2.1}$$

and covariance matrix

$$V\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} = \begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix} . \tag{6.2.2}$$

It is desired to estimate μ_y.

A simple random bivariate sample (x_i,y_i), $i = 1, 2, \ldots, n$ is selected, and then an additional independent simple random sample x_i, $i = n+1, n+2, \ldots, n+n_x$ is selected on $\tilde{x}$ only. A preliminary test of $H_o: \mu_y = \mu_x$ versus $H_A: \mu_y \neq \mu_x$ is made. If $H_o: \mu_y = \mu_x$ is accepted, the pooled estimator

$$v = w_1\bar{y}_n + w_2\bar{x}_{n+n_x} \tag{6.2.3}$$

is used, where all sample means in this chapter have been defined in equation (3.3.4). If H_A: $\mu_y \neq \mu_x$ is accepted, the regression estimator

$$w = \bar{y}_n + \beta(\bar{x}_{n+n_x} - \bar{x}_n) \tag{6.2.4}$$

is used. Note that if $\beta = 0$, the above procedure reduces to the estimation procedure previously discussed where the information on μ_x is incorporated only into the pooled estimator.

The estimation scheme is represented in Figure 6.1.

$$(x_1, y_1)$$
$$(x_2, y_2)$$
$$\vdots$$
$$(x_n, y_n)$$

$$\begin{bmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_{n+n_x} \end{bmatrix}$$

Preliminary Test

H_o: $\mu_y = \mu_x$ versus H_A: $\mu_y \neq \mu_x$

Accept H_o: $\hat{\mu}_{yR} = w_1\bar{y}_n + w_2\bar{x}_{n+n_x} = v$

Accept H_A: $\hat{\mu}_{yR} = \bar{y}_n + \beta(\bar{x}_{n+n_x} - \bar{x}_n) = w$

Figure 6.1. Sampling and estimation scheme with regression estimator

C. Mathematical Specification of the Preliminary Test Estimation Procedure with Regression Estimator

For the same reasons as those mentioned in section C of Chapter III, the random variable $(\tilde{x}, \tilde{y})$ is assumed to follow the bivariate normal distribution with mean given in equation (6.2.1) and known covariance matrix given in equation (6.2.2). For the moment, β is assumed tc be some known constant with no further specification. An optimal choice of β is discussed in section H of this chapter. The covariance structure of the sample observations is given in equation (3.3.3). A preliminary investigation of the procedure with unknown covariance matrix is presented in Chapter VIII.

D. Choice of Test Statistic for the Preliminary Test

Either z or z' can be used to perform the preliminary test, where

$$\begin{aligned} z &= \bar{y}_n - \bar{x}_{n+n_x} \\ z' &= \bar{y}_n - \bar{x}_n. \end{aligned} \tag{6.4.1}$$

z is chosen as the test statistic, following an argument similar to the one presented in section D of Chapter III. From equation (3.4.9)

$$V(\tilde{z}) < V(\tilde{z}') \text{ if, and only if, } \rho < \sigma_x/2\sigma_y, \tag{6.4.2}$$

where, from equation (3.4.6), with $n_y = 0$,

$$V(\tilde{z}) = \sigma_z^2 = \sigma_y^2/n + \sigma_x^2/(n+n_x) - 2\rho\sigma_x\sigma_y/(n+n_x). \tag{6.4.3}$$

E. Derivation of Bias of $\hat{\mu}_{yR}$

$E(\tilde{\hat{\mu}}_{yR})$ can be derived by following the approach as presented in Chapter III in detail. In the following derivation only those arguments unique to this chapter are emphasized.

Rewriting $\bar{x}_{n+n_x}$ as

$$\bar{x}_{n+n_x} = \bar{y}_n - z, \tag{6.5.1}$$

then equations (6.2.3) and (6.2.4) can be written as

$$v = \bar{y}_n - w_2 z$$

and

$$w = (\beta + 1)\bar{y}_n - \beta z - \beta\bar{x}_n. \tag{6.5.2}$$

With ξ_α and $h(z)$ as defined in equations (3.7.3) and (3.7.2),

$$\begin{aligned} E(\tilde{\hat{\mu}}_{yR}) &= E(\tilde{w}) + \int_{|z|<\xi_\alpha\sigma_z} \left\{E(\tilde{v} \mid z) - E(\tilde{w} \mid z)\right\} h(z)dz \\ &= \mu_y + \int_{|z|<\xi_\alpha\sigma_z} \left[(\beta-w_2)z + \beta E(\tilde{\bar{x}}_n - \tilde{\bar{y}}_n \mid z)\right] h(z)dz. \end{aligned} \tag{6.5.3}$$

Now, $E(\tilde{\bar{y}}_n \mid z)$ and $V(\tilde{\bar{y}}_n \mid z)$ are given in equations (3.7.15) and (3.7.16) as

$$\begin{aligned} E(\tilde{\bar{y}}_n \mid z) &= \mu_y + k_1(z - \Delta)/(k_1+k_2) \\ V(\tilde{\bar{y}}_n \mid z) &= \sigma_y^2/n - k_1^2/(k_1+k_2) \end{aligned} \tag{6.5.4}$$

where now, since $n_y = 0$ in this chapter,

$$\begin{aligned}
\Delta &= \mu_y - \mu_x \\
k_1 &= \sigma_y^2/n - \rho\sigma_x\sigma_y/(n+n_x) \\
k_2 &= \sigma_x^2/(n+n_x) - \rho\sigma_x\sigma_y/(n+n_x) \\
k_1 + k_2 &= \sigma_z^2.
\end{aligned} \tag{6.5.5}$$

Similarly,

$$\begin{aligned}
E(\tilde{\bar{x}}_n \mid z) &= \mu_x - k_3(z - \Delta)/(k_1+k_2) \\
V(\tilde{\bar{x}}_n \mid z) &= \sigma_x^2/n - k_3^2/(k_1+k_2),
\end{aligned} \tag{6.5.6}$$

where

$$k_3 = \sigma_x^2/(n+n_x) - \rho\sigma_x\sigma_y/n. \tag{6.5.7}$$

Substituting equations (6.5.4) and (6.5.6) into $E(\tilde{\hat{\mu}}_{yR})$ in equation (6.5.3) yields

$$E(\tilde{\hat{\mu}}_{yR}) = \mu_y - \int_{|z|<\xi_\alpha\sigma_z} \left[\bar{w}_2 z - \beta(z-\Delta)(k_2-k_3)/(k_1+k_2)\right] \; h(z)dz \tag{6.5.8}$$

or, using the transformation $t = (z-\Delta)/\sigma_z$,

$$E(\tilde{\hat{\mu}}_{yR}) = \mu_y - \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[\bar{w}_2\sigma_z(t+\delta) - \beta(k_2-k_3)t/\sigma_z\right] \; \phi(t)\; dt. \tag{6.5.9}$$

$\phi(t)$ is defined in equation (3.7.4), and

$$\begin{aligned}
k_2 - k_3 &= n_x n^{-1}(n+n_x)^{-1} \; \rho\sigma_x\sigma_y \\
\delta &= \Delta/\sigma_z .
\end{aligned} \tag{6.5.10}$$

The bias of the estimator $\hat{\mu}_{yR}$, as a function of δ, is

$$B_R(\delta) = -\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[w_2\sigma_z(t+\delta) - \beta(k_2-k_3)t/\sigma_z\right] \phi(t)\, dt. \tag{6.5.11}$$

$B_R(\delta)$ can be calculated with the aid of tables of the N(0, 1) density function $\phi(t)$ and the N(0, 1) cumulative distribution function $\Phi(t)$ since

$$\begin{aligned} B_R(\delta) = &\left\{w_2\sigma_z - \beta(k_2-k_3)/\sigma_z\right\} \left\{\phi(\xi_\alpha-\delta) - \phi(-\xi_\alpha-\delta)\right\} \\ &- w_2\sigma_z\delta \left\{\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right\}. \end{aligned} \tag{6.5.12}$$

Note that

$$B_R(0) = 0 \tag{6.5.13}$$

when β and w_2 are arbitrary constants. Also,

$$B_R(-\delta) = -B_R(\delta) \tag{6.5.14}$$

and thus by Lemma 3.4 of Chapter III,

$$\lim_{|\delta|\to\infty} B_R(\delta) = 0 \tag{6.5.15}$$

for arbitrary constants β and w_2 which are not functions of δ.

Finally, if $\beta = 0$, then equation (6.5.11) is the same as equation (3.7.23) in Chapter III with $m_x = m_y = n_y = 0$.

F. Derivation of Mean Square Error of $\hat{\mu}_{yR}$

Proceeding as in the previous section, $E(\tilde{\hat{\mu}}^2_{yR})$ is easily obtained as

$$E(\tilde{\hat{\mu}}^2_{yR}) = V(\tilde{w}) + \mu_y^2 + \int_{|z|<\xi_\alpha\sigma_z} E(\tilde{v}^2 - \tilde{w}^2 \mid z)\, h(z)\, dz. \tag{6.6.1}$$

Now,

$$v^2 - w^2 = -(\beta^2+2\beta)\bar{y}_n^2 + (w_2^2 - \beta^2)\, z^2$$
$$+ 2\left[\beta(\beta+1) - w_2\right]\, \bar{y}_n z - \beta^2 \bar{x}_n^2 - \beta^2 \bar{x}_n z$$
$$+ 2\beta(\beta+1)\, \bar{x}_n \bar{y}_n. \qquad (6.6.2)$$

From equations (6.4.1) and (3.3.4)

$$\bar{x}_n = (\bar{y}_n - z)\left[1+n_x/n\right] - n_x \bar{x}_{n_x}/n, \qquad (6.6.3)$$

and substituting this for $\bar{x}_n$ in the last term of equation (6.6.2) yields

$$v^2 - w^2 = \left[\beta^2 + 2\beta(\beta+1)n_x/n\right] \bar{y}_n^2 - 2\left[w_2 + \beta(\beta+1)n_x/n\right] \bar{y}_n z$$
$$+ (w_2^2 - \beta^2)z^2 - \beta^2 \bar{x}_n^2 - \beta^2 \bar{x}_n z - 2\beta(\beta+1)n_x \bar{x}_{n_x} \bar{y}_n/n. \qquad (6.6.4)$$

Using equations (6.5.4) and (6.5.6), every term in $E(\tilde{v}^2 - \tilde{w}^2 \mid z)$ can be evaluated except $E(\tilde{\bar{x}}_{n_x} \tilde{\bar{y}}_n \mid z)$. Although $\tilde{\bar{x}}_{n_x}$ and $\tilde{\bar{y}}_n$ are independent, neither $\tilde{\bar{x}}_{n_x}$ nor $\tilde{\bar{y}}_n$ is independent of $\tilde{z}$. Thus, it is necessary to derive $E(\tilde{\bar{x}}_{n_x} \tilde{\bar{y}}_n \mid z)$.

By Lemma 3.1, the joint distribution of $\tilde{\bar{x}}_{n_x}$, $\tilde{\bar{y}}_n$, and $\tilde{z}$ is trivariate normal. Using Theorem 2.5.1 in Anderson (1958), the conditional joint distribution of $(\tilde{\bar{x}}_{n_x}, \tilde{\bar{y}}_n)$, given z, is bivariate normal. The conditional expectation and covariance matrix of $(\tilde{\bar{x}}_{n_x}, \tilde{\bar{y}}_n)$, given z, can be found from Theorem 2.5.1 of Anderson (1958) after some elementary matrix algebra.

Since

$$E(\tilde{\bar{x}}_{n_x} \tilde{\bar{y}}_n \mid z) = \text{cov}(\tilde{\bar{x}}_{n_x}, \tilde{\bar{y}}_n \mid z) + E(\tilde{\bar{x}}_{n_x} \mid z)E(\tilde{\bar{y}}_n \mid z), \qquad (6.6.5)$$

then $E(\tilde{\bar{x}}_{n_x} \tilde{\bar{y}}_n \mid z)$ is easily found to be

$$E(\tilde{\bar{x}}_{n_x} \tilde{\bar{y}}_n \mid z) = k_1\sigma_x^2 \Big/ \left\{ \sigma_z^2(n+n_x) \right\} + \left\{ \mu_y + \frac{k_1(z-\Delta)}{\sigma_z^2} \right\} \left\{ \mu_y - \Delta - \frac{\sigma_x^2(z-\Delta)}{\sigma_z^2(n+n_x)} \right\}. \qquad (6.6.6)$$

With $(v^2 - w^2)$ as in equation (6.6.4), $E(\tilde{v}^2 - \tilde{w}^2 \mid z)$ can now be evaluated, and then equation (6.6.1) yields

$$E(\tilde{\hat{\mu}}_{yR}^2) = \mu_y^2 + V(\tilde{w}) + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[K_2t^2 + K_1t + K_0\right] \phi(t)\, dt$$

$$- 2\mu_y \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[w_2(t+\delta)\sigma_z - \beta(k_2-k_3)t/\sigma_z\right] \phi(t)\, dt, \qquad (6.6.7)$$

where

$$K_2 = w_2(w_2k_2-w_1k_1-k_1) + 2\beta k_1(k_2-k_3)/(k_1+k_2) - \beta^2(k_2-k_3)^2/(k_1+k_2), \qquad (6.6.8)$$

$$K_1 = 2w_2\delta(w_2k_2-w_1k_1), \qquad (6.6.9)$$

and

$$K_0 = w_2^2\sigma_z^2\delta^2 + 2\beta k_2(k_2-k_3)/(k_1+k_2) + \beta^2(k_2-k_3)^2/(k_1+k_2) - \beta^2 n_x\sigma_x^2/\left[n(n+n_x)\right]. \qquad (6.6.10)$$

Expressing the mean square error as a function of δ as in Chapter III yields

$$MSE_R(\delta) = V(\tilde{w}) + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[K_2 t^2 + K_1 t + K_o\right] \phi(t)\, dt. \tag{6.6.11}$$

It is easily shown that

$$V(\tilde{w}) = \sigma_y^2/n - 2\beta\rho\sigma_x\sigma_y n_x/\left[n(n+n_x)\right] + \beta^2\sigma_x^2 n_x/\left[n(n+n_x)\right] . \tag{6.6.12}$$

$MSE_R(\delta)$ can be calculated with the aid of standard tables of $\phi(t)$ and $\Phi(t)$ since

$$\begin{aligned} MSE_R(\delta) = V(\tilde{w}) &+ \left[K_2+K_o\right]\left[\Phi(\xi_\alpha-\delta) - \Phi(-\xi_\alpha-\delta)\right] \\ &- \left[K_1+K_2(\xi_\alpha-\delta)\right]\phi(\xi_\alpha-\delta) \\ &+ \left[K_1-K_2(\xi_\alpha+\delta)\right]\phi(-\xi_\alpha-\delta). \end{aligned} \tag{6.6.13}$$

By the use of Lemma 3.4, it is easily shown that

$$\lim_{|\delta|\to\infty} MSE_R(\delta) = 0 \tag{6.6.14}$$

when β and w_2 are constants independent of δ. Note that if $\beta = 0$ in equation (6.6.11) for $MSE_R(\delta)$, the same formula is obtained as that given for $MSE(\delta)$ in Chapter III when $m_x = m_y = n_y = 0$.

G. Optimum Choice for the Weights w_1 and w_2

$MSE_R(\delta)$ can be considered as a function of β and w_2, since these variables can be specified by the experimenter. It is desirable, therefore, to choose β and w_2 such that $MSE_R(\delta)$ is minimized. The usual procedure for finding the extrema of a function of two variables can be followed, but some simplifications occur. First, from equations (6.6.8) through (6.6.11), $MSE_R(\delta)$ can be written as

$$MSE_R(\delta) = F_1(w_2) + F_2(\beta) + F_3, \tag{6.7.1}$$

where $F_1(w_2)$ contains terms with w_2 and not β, $F_2(\beta)$ contains terms with β and not w_2, and F_3 contains terms with neither w_2 nor β. Using equation (6.7.1) the second mixed partial derivative of $MSE_R(\delta)$, i.e. $\frac{\partial^2}{\partial\beta\partial w_2} MSE_R(\delta)$, is zero. Further, the two equations

$$\begin{aligned}
\frac{\partial}{\partial w_2} MSE_R(\delta) \Big|_{w_2=w_{20}} &= 0 = \frac{d}{dw_2} F_1(w_2) \Big|_{w_2=w_{20}} \\
\frac{\partial}{\partial\beta} MSE_R(\delta) \Big|_{\beta=\beta_o} &= 0 = \frac{d}{d\beta} F_2(\beta) \Big|_{\beta=\beta_o}
\end{aligned} \tag{6.7.2}$$

can be solved separately for β_o and w_{20} rather than simultaneously since $\frac{\partial}{\partial w_2} MSE_R(\delta)$ does not contain β and, likewise, $\frac{\partial}{\partial\beta} MSE_R(\delta)$ does not contain w_2. Finally, the criterion for a minimum of $MSE_R(\delta)$ is satisfied if the solutions w_{20} and B_o of equations (6.7.2) satisfy

$$\begin{aligned}
\frac{d^2F_1(w_2)}{dw_2^2} \Bigg|_{w_2=w_{20}} &> 0 \\
\frac{d^2F_2(\beta)}{d\beta^2} \Bigg|_{\beta=\beta_o} &> 0 .
\end{aligned} \tag{6.7.3}$$

Thus, the values w_{20} and β_o which minimize $F_1(w_2)$ and $F_2(\beta)$ also minimize $MSE_R(\delta)$.

The optimum weight w_{20} is now derived. Letting $w_1 = 1 - w_2$, the first partial derivative of $MSE_R(\delta)$ with respect to w_2 or the first derivative of $F_1(w_2)$ with respect to w_2 is

$$\frac{d}{dw_2} F_1(w_2) = 2 \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[\begin{array}{c} \left[w_2(k_1+k_2)-k_1\right] t^2 + \delta t \left[2w_2(k_1+k_2)-k_1\right] \\ + w_2\delta^2(k_1+k_2) \end{array} \right] \phi(t)\,dt. \tag{6.7.4}$$

Setting the first derivative equal to zero and solving for w_{20} yields the solution

$$w_{20} = k_1 A(\delta) / \left[(k_1+k_2)C(\delta)\right] , \tag{6.7.5}$$

where $A(\delta)$ and $C(\delta)$ are defined in equations (4.4.3) and (4.4.4) of Chapter IV. This is the same optimum solution which was obtained in equation (4.4.2), but note that k_1 and k_2 have slightly different definitions in Chapters IV and VI since $n_y = 0$ in Chapter IV. (See (4.2.3) and (6.5.5).)

From the discussion of $A(\delta)$ and $C(\delta)$ in Chapter IV, $A(-\delta) = A(\delta)$ and $C(\delta) = C(-\delta)$. From the investigation in Chapter IV of $\lim_{\delta\to\infty} A(\delta)/C(\delta)$, the limit of w_{20} in equation (6.7.5) as δ approaches ∞ is easily found to be

$$\lim_{\delta\to\infty} w_{20} = \left\{ \begin{array}{ll} -\infty & \text{if } k_1 > 0 \\ \infty & \text{if } k_1 < 0 \end{array} \right\} . \tag{6.7.6}$$

See section D of Chapter IV for comments on the infinite limit.

The second derivative of $F_1(w_2)$ with respect to w_2 is

$$\frac{d^2}{dw_2^2} F_1(w_2) = 2\sigma_z^2 \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (t+\delta)^2 \phi(t)\, dt. \tag{6.7.7}$$

Since the second derivative is positive and is not a function of w_2, then w_{20} as given in equation (6.7.5) gives a minimum value for $F_1(w_2)$ or $MSE_R(w_2)$ with respect to w_2.

Substituting the optimum value w_{20} into $MSE_R(\delta)$ and denoting this by $MSE_R(w_{20})$ yields

$$\begin{aligned} MSE_R(w_{20}) &= \frac{\sigma_y^2}{n} - \frac{k_1^2 A^2(\delta)}{(k_1+k_2)C(\delta)} \\ &+ 2\beta\left[\frac{-\rho\sigma_x\sigma_y n_x}{n(n+n_x)} + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left\{\frac{k_1(k_2-k_3)t^2}{\sigma_z^2} + \frac{k_2(k_2-k_3)}{\sigma_z^2}\right\} \phi(t)dt\right] \\ &+ \beta^2\left[\frac{\sigma_x^2 n_x}{n(n+n_x)} + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left\{\frac{-t^2(k_2-k_3)^2}{\sigma_z^2} + \frac{(k_2-k_3)^2}{\sigma_z^2} - \frac{n_x\sigma_x^2}{n(n+n_x)}\right\} \phi(t)dt\right] \end{aligned} \tag{6.7.8}$$

In section D of Chapter IV it was shown that, for ξ_α finite,

$$\lim_{\delta\to\infty} A^2(\delta)/C(\delta) = 0. \tag{6.7.9}$$

Thus, if β is a constant independent of δ, then

$$\lim_{\delta\to\infty} MSE_R(w_{20}) = V(\tilde{w}),\ \xi_\alpha \text{ finite}. \tag{6.7.10}$$

Of course, w_{20} can not be determined exactly, since Δ (and thus δ) is unknown. See comments in section D of Chapter IV for approximations to w_{20}. If nothing is known about δ, then for simplicity w_{20} can be evaluated at $\delta = 0$, which results in

$$\begin{aligned} w_2 &= k_1/(k_1+k_2) \\ w_1 &= k_2/(k_1+k_2), \end{aligned} \tag{6.7.11}$$

where k_1 and k_2 are given in equation (6.5.5). These are also the weights that minimize the variance of the pooled estimator $v = w_1\bar{y}_n + w_2\bar{x}_{n+n_x}$. Further, letting $\delta = 0$ in order to evaluate w_{20} can be thought of as an approximation obtained by putting a $N(0, a^2)$ prior distribution on $\tilde{\Delta}$ and then approximating $\delta = \Delta/\sigma_z$ by the prior mean of $\tilde{\Delta}/\sigma_z$.

H. Optimum Choice for the Regression Coefficient β

If the regression estimator $w = \bar{y}_n + \beta(\bar{x}_{n+n_x} - \bar{x}_n)$ is used all the time, then the choice of β which minimizes $V(\tilde{w})$ as given in equation (6.6.12) is $B = \rho\sigma_y/\sigma_x$, the usual regression coefficient as discussed in Cochran (1963). However, the constant $\rho\sigma_y/\sigma_x$ may not be the optimum choice for β in the preliminary test estimation scheme with mean square error as given in equation (6.6.11). Thus, it is necessary to find the value of β which minimizes $MSE_R(\delta)$ in the preliminary test estimation scheme.

The first derivative of $F_2(\beta)$ in equation (6.7.1) with respect to β is

$$\frac{d}{d\beta} F_2(\beta) = -\frac{2\rho\sigma_x\sigma_y n_x}{n(n+n_x)} + \frac{2\beta\sigma_x^2 n_x}{n(n+n_x)}$$

$$+ \frac{2(k_2-k_3)}{(k_1+k_2)} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[k_1 - \beta(k_2-k_3)\right] t^2 \phi(t)\, dt$$

$$+ 2\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left\{\frac{k_2(k_2-k_3)}{(k_1+k_2)} + \frac{\beta(k_2-k_3)^2}{(k_1+k_2)} - \frac{\beta n_x\sigma_x^2}{n(n+n_x)}\right\} \phi(t)\, dt$$

(6.8.1)

Also, this same derivative can be obtained by differentiating $MSE_R(w_{20})$ in equation (6.7.8) with respect to β. Setting the above derivative equal to zero and solving for the optimum value β_o yields

$$\beta_o = \frac{\dfrac{\rho\sigma_x\sigma_y n_x}{n(n+n_x)} - W(\delta)}{\dfrac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)}, \qquad (6.8.2)$$

where

$$W(\delta) = \frac{(k_2-k_3)}{(k_1+k_2)} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (k_1 t^2 + k_2)\, \phi(t)\, dt, \qquad (6.8.3)$$

and

$$Z(\delta) = \frac{(k_2-k_3)^2}{(k_1+k_2)} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (t^2-1)\, \phi(t)\, dt + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \frac{n_x\sigma_x^2}{n(n+n_x)}\, \phi(t)\, dt.$$

(6.8.4)

By Lemma 3.4 of Chapter III

$$\lim_{\delta\to\infty} W(\delta) = \lim_{\delta\to\infty} Z(\delta) = 0, \tag{6.8.5}$$

and thus

$$\lim_{\delta\to\infty} \beta_o = \rho\sigma_y/\sigma_x. \tag{6.8.6}$$

The second derivative of $F_2(\beta)$ with respect to β is

$$\frac{d^2\left[F_2(\beta)\right]}{d\beta^2} = \frac{2\sigma_x^2 n_x}{n(n+n_x)} + 2\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left\{\frac{(k_2-k_3)^2(1-t^2)}{(k_1+k_2)} - \frac{n_x\sigma_x^2}{n(n+n_x)}\right\}\phi(t)\,dt. \tag{6.8.7}$$

Note that the second derivative of $F_2(\beta)$ does not depend upon β. In order to show that β_o yields a minimum for $F_2(\beta)$, it is necessary to show that $\frac{d^2}{d\beta^2}F_2(\beta) > 0$. Since $V(\tilde{x} - \tilde{y}) > 0$ and $|\rho| < 1$, then

$$n(\sigma_y^2 + \sigma_x^2 - 2\rho\sigma_x\sigma_y) + n_x\sigma_y^2(1-\rho^2) > 0. \tag{6.8.8}$$

Multiplying both sides of the inequality by $n_x\sigma_x^2/\left[n^2(n+n_x)^2\right]$ and rearranging terms yields

$$\frac{n_x\sigma_x^2}{n(n+n_x)}\left\{\frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n+n_x} - \frac{2\rho\sigma_x\sigma_y}{n+n_x}\right\} - \frac{n_x^2\sigma_x^2\sigma_y^2\rho^2}{n^2(n+n_x)^2} > 0. \tag{6.8.9}$$

From equations (6.5.5) and (6.5.10), equation (6.8.9) becomes

$$n_x\sigma_x^2/\left[n(n+n_x)\right] - (k_2-k_3)^2/(k_1+k_2) > 0, \tag{6.8.10}$$

for $(k_1+k_2) = \sigma_z^2 \neq 0$. Thus, in equation (6.8.4) for $Z(\delta)$,

$$\frac{(t^2-1)(k_2-k_3)^2}{(k_1+k_2)} + \frac{n_x\sigma_x^2}{n(n+n_x)} > 0 \text{ for all } t. \tag{6.8.11}$$

Since $\phi(t) > 0$ for all t, then

$$0 < \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left\{ \frac{(t^2-1)(k_2-k_3)^2}{(k_1+k_2)} + \frac{n_x\sigma_x^2}{n(n+n_x)} \right\} \phi(t)\, dt$$

$$= Z(\delta) < \int_{-\infty}^{\infty} \left\{ \frac{(t^2-1)(k_2-k_3)^2}{(k_1+k_2)} + \frac{n_x\sigma_x^2}{n(n+n_x)} \right\} \phi(t)\, dt$$

$$= E\left[\frac{(\tilde{T}^2-1)(k_2-k_3)^2}{(k_1+k_2)} + \frac{n_x\sigma_x^2}{n(n+n_x)} \right]$$

$$= \frac{n_x\sigma_x^2}{n(n+n_x)}\,, \tag{6.8.12}$$

where $\tilde{T}$ is distributed $N(0, 1)$. Thus

$$0 < Z(\delta) < n_x\sigma_x^2 / \left[n(n+n_x)\right]\,, \tag{6.8.13}$$

and

$$\frac{d^2}{d\beta^2}\left[MSE_R(\delta)\right] = 2\left\{\frac{n_x\sigma_x^2}{n(n+n_x)} - Z(\delta)\right\} > 0. \tag{6.8.14}$$

Since $\frac{d^2}{d\beta^2}\left[F_2(\beta)\right]$ does not depend upon β, and since it is positive, then β_o minimizes $F_2(\beta)$ with respect to β.

Taking equation (6.7.8) for $MSE_R(w_{20})$, and letting $\beta = \beta_o$ as in equation (6.8.2), the minimum mean square error with $\beta = \beta_o$ and $w_2 = w_{20}$, denoted by $MSE_R(w_{20}, \beta_0)$, is

$$MSE_R(w_{20}, \beta_0) = \frac{\sigma_y^2}{n} - \frac{k_1^2 A^2(\delta)}{(k_1+k_2)C(\delta)}$$

$$- \frac{\left\{\frac{\rho\sigma_x\sigma_y n_x}{n(n+n_x)} - W(\delta)\right\}^2}{\left\{\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)\right\}} . \qquad (6.8.15)$$

Using equations (6.7.9) and (6.8.5),

$$\lim_{\delta \longrightarrow \infty} MSE_R(w_{20}, \beta_0) = \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2\sigma_y^2 n_x}{n(n+n_x)}\right\}. \qquad (6.8.16)$$

If $\beta = 0$ in equation (6.7.8), then

$$MSE_R(w_{20}) = \frac{\sigma_y^2}{n} - \frac{k_1^2 A^2(\delta)}{(k_1+k_2)C(\delta)} . \qquad (6.8.17)$$

Note that $MSE_R(w_{20})$ above is always less than σ_y^2/n since $C(\delta) > 0$ and $(k_1+k_2) > 0$. Thus, with the optimum weight w_{20} in equation (6.7.5), the preliminary test scheme with $\beta = 0$ has smaller mean square error than the procedure which uses $\hat{\mu}_y = \bar{y}_n$ all the time. Now, consider $MSE_R(w_{20}, \beta_0)$ in equation (6.8.15). Since $\sigma_x^2 n_x / [n(n+n_x)] > Z(\delta)$, then

$$MSE_R(w_{20}, \beta_0) < MSE_R(w_{20}) < \sigma_y^2/n. \qquad (6.8.18)$$

Thus, the preliminary test scheme with optimum regression coefficient β_0 and optimum weight w_{20} always has smaller mean square error than the preliminary test scheme with $\beta = 0$ and optimum weight w_{20}.

Consider now the preliminary test procedure with $\beta = 0$ and $w_2 = k_1/(k_1+k_2)$. The mean square error of this procedure is then given by equation (6.6.11) as

$$\mathrm{MSE}(\delta) = \frac{\sigma_y^2}{n} + \frac{k_1^2}{\sigma_z^2} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (\delta^2-t^2)\, \phi(t)\, dt. \tag{6.8.19}$$

Consider again equation (6.6.11) with $w_2 = k_1/(k_1+k_2)$, but with β now some arbitrary constant. Then the mean square error is given as

$$\begin{aligned}\mathrm{MSE}_R(\delta) &= \frac{\sigma_y^2}{n} - 2\beta(k_2-k_3) + \frac{\beta^2\sigma_x^2 n_x}{n(n+n_x)} \\ &+ \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left[-\frac{k_1^2}{\sigma_z^2} + \frac{2\beta k_1(k_2-k_3)}{\sigma_z^2} - \frac{\beta^2(k_2-k_3)^2}{\sigma_z^2}\right] t^2\phi(t)\, dt \\ &+ \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left[\frac{k_1^2\delta^2}{\sigma_z^2} + \frac{2\beta k_2(k_2-k_3)}{\sigma_z^2} + \frac{\beta^2(k_2-k_3)^2}{\sigma_z^2} - \frac{\beta^2 n_x\sigma_x^2}{n(n+n_x)}\right]\phi(t)dt.\end{aligned} \tag{6.8.20}$$

The difference in the mean square error between the preliminary test scheme with $\beta = 0$ and the preliminary test scheme with regression estimator when $w_2 = k_1/(k_1+k_2)$ for both is obtained from equations (6.8.19) and (6.8.20), after some algebra, as

$$MSE(\delta) - MSE_R(\delta) = 2\beta\left[(k_2-k_3)-W(\delta)\right] - \beta^2\left[\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)\right] \quad (6.8.21)$$

where $W(\delta)$ and $Z(\delta)$ are defined in equations (6.8.3) and (6.8.4). If the regression scheme is better than the regular preliminary test scheme, then the difference in equation (6.8.21) should be positive. If now equation (6.8.2) is evaluated at some value of δ, say δ^*, to yield a value for β, then equation (6.8.21) becomes

$$MSE(\delta) - MSE_R(\delta) = \frac{2\left[(k_2-k_3)-W(\delta)\right]\left[(k_2-k_3)-W(\delta^*)\right]}{\left[\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta^*)\right]} - \frac{\left[(k_2-k_3)-W(\delta^*)\right]^2\left[\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)\right]}{\left[\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta^*)\right]^2} . \quad (6.8.22)$$

Now, $\left\{\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)\right\} > 0$ for any δ by equation (6.8.13). Thus, equation (6.8.22) implies

$MSE(\delta) - MSE_R(\delta) \geq 0$ if, and only if,

$$\frac{\left\{\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta^*)\right\}\left\{(k_2-k_3) - W(\delta)\right\}}{\left\{\frac{\sigma_x^2 n_x}{n(n+n_x)} - Z(\delta)\right\}\left\{(k_2-k_3) - W(\delta^*)\right\}} \geq \frac{1}{2} . \quad (6.8.23)$$

Obviously, if δ^* equals the population value of δ, then inequality (6.8.23) is satisfied. From inequality (6.8.23) it can be concluded that, with $w_2 = k_1/(k_1+k_2)$, the mean square error for the preliminary test with regression estimator will be less than the mean square error for the preliminary test without regression whenever δ^* is near the population value of δ. This is a sufficient condition only. A further analysis of inequality (6.8.23) may reveal weaker sufficient conditions.

In addition to evaluating equation (6.8.2) at $\delta = \delta^*$ to obtain a value for β, other approximations may be made as discussed in Chapter IV. If nothing is known about δ, then it seems best, intuitively, to use the regular regression coefficient $\beta = \rho\sigma_y/\sigma_y$. For the purpose of comparison in Chapter VIII, the preliminary test scheme with regression estimator is evaluated for $\beta = \rho\sigma_y/\sigma_x$ and $w_2 = k_1/(k_1+k_2)$ in the next section.

I. Bias and Mean Square Error when $\beta = \rho\sigma_y/\sigma_x$ and $w_2 = k_1/(k_1+k_2)$

From equation (6.5.11), with $\beta = \rho\sigma_y/\sigma_x$ and $w_2 = k_1/(k_1+k_2)$, the bias is obtained as

$$B_R(\delta) = -\frac{1}{\sigma_z}\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left[k_1\delta + t\left\{k_1 - \frac{\rho^2\sigma_y^2 n_x}{n(n+n_x)}\right\}\right]\phi(t)\,dt. \tag{6.9.1}$$

As was noted earlier, $B_R(0) = 0$ and $\lim_{\delta\to\infty} B_R(\delta) = 0$.

From equation (6.6.11) the mean square error is obtained as

$$\mathrm{MSE}_R(\delta) = \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{(n+n_x)}\right\} + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} \left[K_2 t^2 + K_0\right] \phi(t)\, dt, \qquad (6.9.2)$$

where now, with $\beta = \rho\sigma_y/\sigma_x$ and $w_2 = k_1/(k_1+k_2)$,

$$K_2 = -\frac{1}{\sigma_z^2}\left\{k_1 - \frac{\rho^2\sigma_y^2 n_x}{n(n+n_x)}\right\}^2 \qquad (6.9.3)$$

and

$$K_0 = \frac{1}{\sigma_z^2}\left\{k_1^2\delta^2 + \frac{\rho^2\sigma_y^2 n_x(k_2-k_1)}{n(n+n_x)} + \frac{\rho^4\sigma_y^4 n_x^2}{n^2(n+n_x)^2}\right\}. \qquad (6.9.4)$$

Using Lemma 3.4 it follows easily that

$$\lim_{\delta\to\infty} \mathrm{MSE}_R(\delta) = \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{(n+n_x)}\right\}. \qquad (6.9.5)$$

In sampling theory, the usual variance of a regression estimator (neglecting correction factors) is $\sigma_y^2(1-\rho^2)/n$ when the population mean μ_x is known. μ_x is not assumed to be known in this discussion and thus $\frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{(n+n_x)}\right\}$ is obtained as the variance of the regression estimator. Note, however, that

$$\lim_{n_x\to\infty} \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{(n+n_x)}\right\} = \frac{\sigma_y^2(1-\rho^2)}{n}. \qquad (6.9.6)$$

In Chapter VIII comparisons of the preliminary test scheme, the preliminary test scheme with regression estimator, and the regression estimator are made. These comparisons use $\beta = \rho\sigma_y/\sigma_x$ and $w_2 = k_1/(k_1+k_2)$, and hence equations (6.9.1) and (6.9.2) through (6.9.4) will be used for bias and mean square error of the preliminary test scheme with regression estimator.

VII. POOLED BAYESIAN ESTIMATION FOR THE MEAN OF A NORMAL POPULATION

A. Introduction

1. Bayesian theory

In Bayesian statistical theory the random variable $\tilde{y}$ has a distribution conditional upon the parameters $\theta_1, \theta_2, \ldots, \theta_p$, i.e. $f(y \mid \theta_1, \theta_2, \ldots, \theta_p)$. A joint prior probability distribution $g(\theta_1, \theta_2, \ldots, \theta_p)$ is assigned to the unknown parameters. Observations y_i, $i = 1, 2, \ldots, n$, on $\tilde{y}$ are collected, and then the posterior distribution of the parameters, given the observed data, is obtained, i.e. $h(\theta_1, \theta_2, \ldots, \theta_p \mid y_1, y_2, \ldots, y_n)$.

If interest is on a few of several unknown parameters, then the marginal distributions of particular interest are obtained from the joint posterior distribution of all the parameters. In the case of nuisance parameters, note that this procedure consists of putting a prior distribution on the nuisance parameters and then, in effect, "integrating them out." Thus, the posterior distribution of interest does not involve any of the nuisance parameters.

The most commonly used Bayesian estimator is the mean of the posterior distribution. It is easily shown that the posterior mean minimizes the expectation of a quadratic loss function, where the expectation is taken with respect to the posterior distribution of the parameters.

2. Topics investigated in this chapter

Discussed in this chapter is the application of Bayesian estimation theory to the problem of pooling means under the assumption of various prior distributions on $\tilde{\mu}_y$ and $\tilde{\mu}_x$ or on $\tilde{\mu}_y$ and $\tilde{\Delta} = \tilde{\mu}_y - \tilde{\mu}_x$. Some particular aspects of Bayesian theory, such as precise measurement, are discussed in section B. In section C a simple example of Bayesian inference is given to illustrate the method.

Section D defines the distribution for the "observed" random variables under consideration and states the sampling plan to be used. Sufficient statistics in the Bayesian sense are defined.

Section E derives the estimator of μ_y when the prior distribution on $(\tilde{\mu}_y, \tilde{\mu}_x)$ is relatively constant.

In section F a bivariate normal prior distribution is assigned to $\tilde{\mu}_y$ and $\tilde{\mu}_x$, and the estimator of μ_y is derived. The most general results are presented here, and most other sections in Chapter VII are special cases of section F.

Section G is a special case of section F where $\tilde{x}$ and $\tilde{y}$ are independent. The estimator of μ_y is presented, and an optimum sample allocation is derived.

In section H the random variables $\tilde{\mu}_y$ and $\tilde{\Delta}$ are assumed to be independently normally distributed. The results of this section can be derived from section F, and the estimator of μ_y is presented. The important special case where $\tilde{x}$ and $\tilde{y}$ are independent is discussed in detail.

Section I presents the estimator of μ_y when $\tilde{\Delta}$ is normally distributed independently of $\tilde{\mu}_y$, and the principle of precise measurement is applied to $\tilde{\mu}_y$. Included also is some discussion of precise measurement for the case of more than one random variable and parameter. Again, the special case where $\tilde{x}$ and $\tilde{y}$ are independent is discussed separately.

In section J $\tilde{\mu}_y$ and $\tilde{\Delta}$ are independent, where $\tilde{\Delta}$ is normally distributed and $\tilde{\mu}_y$ is uniformly distributed over a specified interval. The estimator of μ_y is more complicated than all preceding ones, and an approximation to the estimator is given. The estimator is discussed for the special case where $\tilde{x}$ and $\tilde{y}$ are independent.

Section K summarizes the types of estimators obtained and compares them in various situations.

B. Some Aspects of Bayesian Statistics

1. The problem of prior distributions

Good (1965, p. 8) gives an interesting definition of a Bayesian as:

> "Several different kinds of Bayesians exist, but it seems to me that the essential defining property of a Bayesian is that he regards it as meaningful to talk about the probability $P[H \mid E]$ of a hypothesis H, given evidence E. Consequently, he will make more use of Bayes' theorem than a non-Bayesian will. Bayes' theorem itself is a trivial consequence of the product axiom of probability, and it is not a belief in this theorem that makes a person a Bayesian. Rather it is a readiness to incorporate intuitive probability into statistical theory and practice."

This readiness to incorporate intuitive probability into statistics via the prior distribution draws heavy criticism from non-Bayesians who say that this method is not objective. Even Bayesians themselves have some difficulty with their method when they have no prior information or intuitive feelings about the form of the prior distribution. Bartholomew (1964, pp. 201-202), after pointing out some advantages of Bayesian theory which were mentioned in the Introduction, says:

> "...the price to be paid for this simplification is the introduction of a prior distribution. To many statisticians this seems to require the abandonment of objectivity in favour of a theory built on the doubtful foundation of individual introspection...the problem is particularly acute if our prior knowledge is vague or nonexistent. However, it would be possible to retain the obvious advantages of the Bayesian approach without sacrificing objectivity if some way of representing ignorance could be found. It is well known that Jeffreys has suggested an invariance principle to achieve this end."

Jeffreys (1961, p. 117) says:

> "A problem of estimation is one where we are given the form of the law, in which certain parameters can be treated as unknown, no special consideration needing to be given to any particular values, and we want the probability distribution of these parameters, given the observations...The essential function of these [invariance principle] rules is to provide a formal way of expressing ignorance of the value of the parameter over the range permitted...Their function is simply to give formal rules, as impersonal as possible, that will enable the theory to begin."

Good (1965, p. 10) speaks further on intuitive probability:

> "An <u>extreme</u> Bayesian believes that every intuitive probability is <u>precise</u>, whereas less extreme Bayesians regard intuitive probabilities as only partially ordered so that each probability merely lies in some interval of values...the less extreme Bayesian...makes judgments of probability inequalities and infers new probability inequalities with the help of the mathematical theory...One is <u>more or less</u> a Bayesian depending on the

precision with which one is prepared to make intuitive probability estimates."

Smith (1961) discusses in more detail the relaxing of the assumption of a precise prior and considers personal probabilities contained only within specified intervals. For example, if a precise prior were placed on $\tilde{\mu}$, then $Pr\left[0 < \tilde{\mu} < 10\right]$ could be evaluated exactly, say $Pr\left[0 < \tilde{\mu} < 10\right] = .36$. However, Smith considers the prior distribution to be specified by a series of statements like $.25 < Pr\left[0 < \tilde{\mu} < 10\right] < .45$. He uses Bayes' theorem as a method of inference and finds that Bayes' theorem remains an equality even though the prior distribution is represented as a series of inequalities.

The above comments by Good (1965), Jeffreys (1961), and Smith (1961) illustrate that Bayesians do indeed recognize situations where a precise prior is not feasible, and they suggest various alternatives.

2. Precise measurement or stable estimation

A prior distribution which is not precise and expresses vague or non-existent information is often called diffuse. Savage (1962) has attacked the problem of diffuse priors with what he calls precise measurement or stable estimation. The principle of precise measurement says that, under certain conditions, the posterior distribution of $\tilde{\mu}_y$ is approximately proportional to the data distribution $f(y \mid \mu_y)$, or to the likelihood function $\prod_{i=1}^{n} f(y_i \mid \mu_y)$ if the experiment results in independent, identically distributed random variables $\tilde{y}_i$. Thus, for all practical purposes, the posterior distribution of $\tilde{\mu}_y$ can be taken as

$$k(\mu_y \mid y_1, y_2, \ldots, y_n) = c \prod_{i=1}^{n} f(y_i \mid \mu_y), \tag{7.2.1}$$

where $c > 0$ is a normalizing constant such that

$$\int_{-\infty}^{\infty} c \prod_{i=1}^{n} f(y_i \mid \mu_y) d\, \mu_y = 1. \tag{7.2.2}$$

This, in effect, removes the problem of even specifying a particular prior distribution $g(\mu_y)$. However, the use of precise measurement depends very much upon the "certain conditions" mentioned previously. Let y_s be a sufficient statistic for the posterior analysis of $\tilde{\mu}_y$, i.e. the posterior distribution of $\tilde{\mu}_y$, given $y_1, y_2, \ldots, y_n$, is the same as the posterior distribution of $\tilde{\mu}_y$, given y_s. Let $k(y_s \mid \mu_y)$ be the density function of $\tilde{y}_s$. Then, the prior distribution of $\tilde{\mu}_y$ must be sufficiently specified to know that it satisfies two properties. (1) For those values of μ_y where $k(y_s \mid \mu_y)$ is large when y_s is the numerical sample result, the prior density $g(\mu_y)$ must be relatively constant. This must be true for any probable sample result y_s. (2) No short interval on the μ_y axis can have a prior probability many times larger than the most improbable interval of the same length selected from a portion of the real line where $g(\mu_y) > 0$. In other words, the prior distribution can have no spikes or regions of extremely concentrated probability. These two requirements are illustrated in Figure 7.1, where the prior density $g(\mu_y)$ and the density $k(y_s \mid \mu_y)$ are not drawn to the same vertical scale.

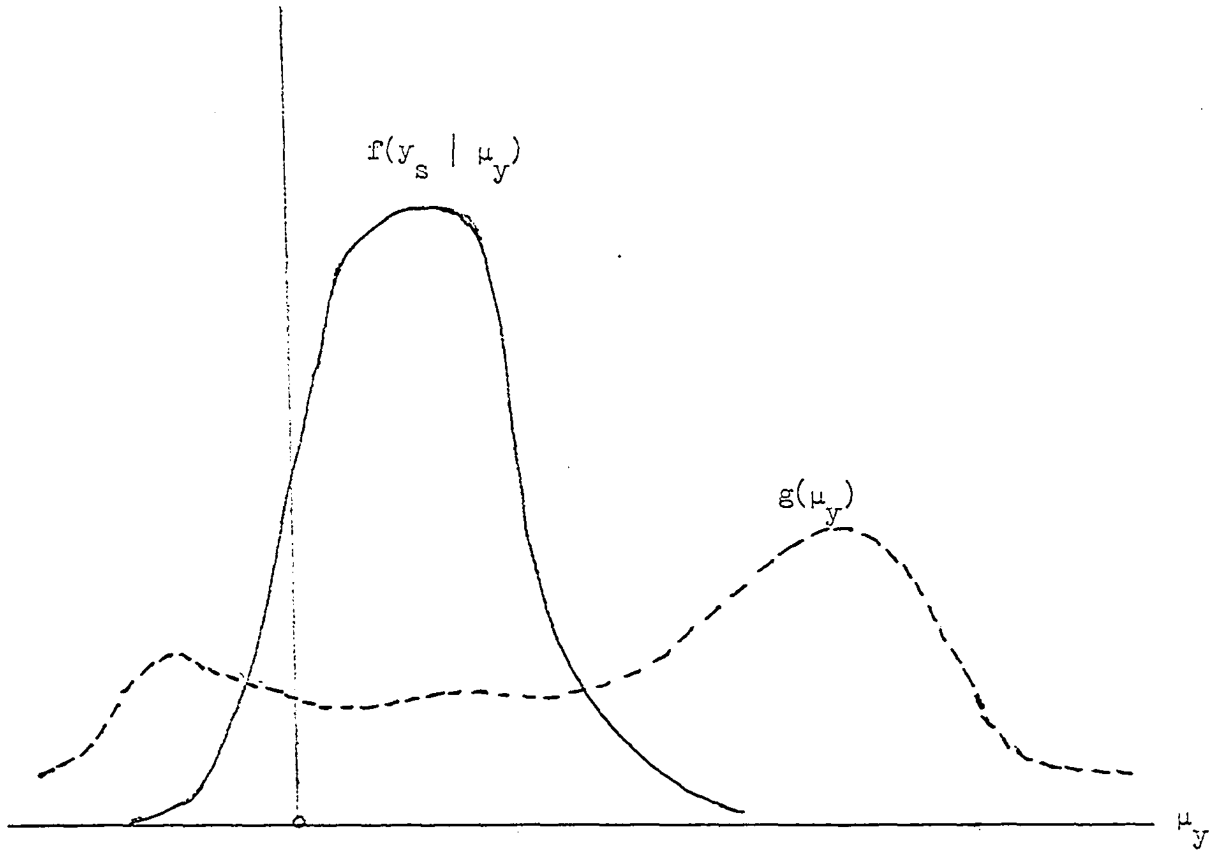

Figure 7.1. A case where precise measurement is applicable

A more thorough discussion of precise measurement can be found in Savage (1962), Edwards et al. (1963), and Savage (1961b). Jeffreys (1961) has obtained a similar result by taking the distribution of $\tilde{\mu}_y$ to be uniform on the real line, sometimes considered nonsensical by mathematical probabilists since the density cannot be normalized. Savage (1962), however, objects to Jeffreys' approach on the grounds that a uniform distribution on the real line is never a realistic prior opinion about a parameter. Edwards et al. (1963, p. 31) give the main point of stable estimation by saying:

> "The method of stable estimation might casually be described as a procedure for ignoring prior opinion since its approximate results are acceptable for a wide range of prior opinions. Actually, far from ignoring prior opinion, stable estimation exploits certain well-defined features of prior opinion and is acceptable insofar as those features are really present."

3. Natural conjugate prior distributions

If the family of probability distributions to which the prior distribution $g(\mu_y)$ belongs is the same as the family to which the posterior distribution $k(\mu_y \mid y_1, y_2, \ldots, y_n)$ belongs, then the family of priors is said to be conjugate to the family of data distributions to which $f(y \mid \mu_y)$ belongs. In this case, the prior and posterior distributions of $\tilde{\mu}_y$ have the same form but different parameters, the parameters of the posterior distribution being a combination of the observed data and the parameters of the prior distribution. The mathematics of Bayesian theory is usually considerably simplified with natural conjugate distributions, and Raiffa and Schlaifer (1961) give the family of natural conjugate distributions for several well known data distributions.

C. An Example of Bayesian Inference

1. Bernoulli distribution

The Bernoulli distribution will be used for an example because it illustrates the Bayesian analysis quite simply and will also give a background for the discussion of the binomial distribution in Chapter IX. The parameter of the distribution is p, the probability of a success, and the conditional distribution of $\tilde{x}$, given $\tilde{p} = p$, is

$$\begin{aligned} f(x \mid p) &= p^{x}(1-p)^{1-x}, \quad x = 0,\ 1 \\ f(x \mid p) &= 0 \quad \text{elsewhere.} \end{aligned} \tag{7.3.1}$$

The Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ is the prior distribution most often used for a proportion, i.e.

$$\begin{aligned} g(p) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1}(1-p)^{\beta-1} \quad 0 \leq p \leq 1 \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{7.3.2}$$

Not only is the Beta distribution the natural conjugate of the Bernoulli distribution, but it has other desirable properties described by Good (1965, p. 17) as:

> "By selecting from the class of beta distributions the statistician can give expression to his initial ideas concerning both the Type II [prior] expectation and the Type II [prior] variance of p, and this is about as much flexibility as is likely to be required in some applications. The distribution has only two parameters, but covers a good variety of unimodal shapes."

With only one observation on $\tilde{x}$, either $\tilde{x} = 0$ or $\tilde{x} = 1$, the joint distribution of the random variables $\tilde{x}$ and $\tilde{p}$ is

$$\begin{aligned} h(x,\ p) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{x+\alpha-1}(1-p)^{\beta-x} \quad \begin{matrix} 0 \leq p \leq 1 \\ x = 0,\ 1 \end{matrix} \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{7.3.3}$$

The integral of h(x, p) with respect to p yields the unconditional distribution of $\tilde{x}$ as

$$f(x) = \int_0^1 h(x,\ p)\ dp = \frac{\Gamma(x+\alpha)\Gamma(\beta-x+1)}{(\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)} \quad \text{for } x = 0,\ 1$$

$$= 0 \quad \text{elsewhere .} \tag{7.3.4}$$

Note that

$$\Pr[\tilde{x} = 0] = f(0) = \beta/(\alpha+\beta)$$
$$\Pr[\tilde{x} = 1] = f(1) = \alpha/(\alpha+\beta), \tag{7.3.5}$$

where $\alpha/(\alpha+\beta)$ is also the mean of the prior distribution $g(p)$. The posterior distribution of $\tilde{p}$ is

$$k(p \mid x) = \frac{h(x,p)}{f(x)} = \frac{\Gamma(\alpha+\beta+1)p^{x+\alpha-1}(1-p)^{\beta-x}}{\Gamma(x+\alpha)\Gamma(\beta-x+1)} \quad \text{for } 0 \leq p \leq 1$$

$$= 0 \quad \text{elsewhere.} \tag{7.3.6}$$

Since $k(p \mid x)$ is a Beta distribution with parameters $(x + \alpha)$ and $(\beta-x+1)$, then

$$E(\tilde{p} \mid x) = (\alpha+x)/(\alpha+\beta+1). \tag{7.3.7}$$

Thus, $\hat{p} = (\alpha+x)/(\alpha+\beta+1)$ is the best estimator for a quadratic loss function.

2. Two unknown parameters

Consider now two Bernoulli random variables $\tilde{x}_1$ and $\tilde{x}_2$ with joint distribution $f(x_1,\ x_2 \mid p_1,\ p_2)$, given $\tilde{p}_1 = p_1$ and $\tilde{p}_2 = p_2$. Let the joint prior distribution be $g(p_1,\ p_2)$. It is desired to estimate p_1, possibly making use of the observations on $\tilde{x}_2$. If $\tilde{x}_1$, given p_1, and $\tilde{x}_2$, given p_2, are independent, and, in addition, $\tilde{p}_1$ and $\tilde{p}_2$ are assumed to be independent, then the posterior mean of $\tilde{p}_1$, i.e. $E[\tilde{p}_1 \mid x_1, x_2]$, will not depend upon x_2. Thus, the observation on $\tilde{x}_2$ contributes nothing to the estimator $\hat{p}_1$. This same result is obtained if the conditional distributions of $\tilde{x}_1$ and $\tilde{x}_2$ are independent as above and the principle of precise measurement is applied to the joint distribution of $\tilde{p}_1$ and $\tilde{p}_2$. This problem arises again in Chapter IX where an adequate prior on $(\tilde{p}_1, \tilde{p}_2)$ is discussed for the pooled Bayesian estimation of a proportion.

The above point has special relevance to pooled estimation. For example, if $\tilde{x}$ and $\tilde{y}$ are bivariate normal with zero correlation and known variances, and the prior distributions on $\tilde{\mu}_x$ and $\tilde{\mu}_y$ are independent, then the Bayesian estimator of μ_y will not be a function of any observations on $\tilde{x}$. Since this dissertation is concerned with pooling observations on $\tilde{x}$ and $\tilde{y}$ for the estimation of μ_y, independent data distributions and independent prior distributions will not be considered.

D. Sampling Scheme and Sufficient Statistics

In the derivations of this chapter the random variable $(\tilde{x}, \tilde{y})$ is distributed bivariate normal with parameters $(\mu_x, \mu_y, \rho, \sigma_x^2, \sigma_y^2)$. As in previous chapters, the case where ρ, σ_x^2, and σ_y^2 are known will be considered first.

The sampling scheme used in this chapter is the special case of the scheme in Figure 3.1 obtained by setting $m_x = m_y = 0$. Thus, it is a one stage sampling scheme with $(n+n_x)$ observations on $\tilde{x}$ and $(n+n_y)$ observations on $\tilde{y}$, where n is the size of the bivariate sample on $(\tilde{x}, \tilde{y})$ and the remaining (n_x+n_y) observations are all independent of each other and the bivariate sample. The notation of equation (3.3.4) is used throughout this chapter.

Once a joint prior distribution on $(\tilde{\mu}_y, \tilde{\mu}_x)$ or $(\tilde{\mu}_y, \tilde{\Delta})$ has been specified, it is necessary to find the posterior distribution of $(\tilde{\mu}_y, \tilde{\mu}_x)$ or $(\tilde{\Delta}, \tilde{\mu}_y)$, given the sample observations $y_1, y_2, \ldots, y_{n+n_y}, x_1, x_2, \ldots, x_{n+n_x}$. As in classical theory it is possible to summarize the sample data into a set of sufficient statistics defined by Raiffa and Schlaifer (1961, pp. x-xi) as:

> "...we define a sufficient statistic as one which leads to the same posterior distribution that would be obtained by use of a 'complete' description of the experimental outcome...we show that this definition implies and is implied by the classical definition of sufficiency in terms of factorability of the joint likelihood of the sample observations..."

Theorem 2.2.2 of Raiffa and Schlaiffer (1961, p. 33) states the factorability criterion for sufficiency in the Bayesian sense.

E. Precise Measurement on $(\tilde{\mu}_x, \tilde{\mu}_y)$

1. <u>Prior distribution on $\tilde{\mu}_x$ and $\tilde{\mu}_y$</u>

In sections E and F two different prior distributions are considered for $(\tilde{\mu}_x, \tilde{\mu}_y)$, and then the estimator of μ_y is derived. In this section the principle of precise measurement is applied to the joint prior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$. In the derivation of the posterior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$

the joint prior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$ is thus considered to be a constant. Justification for this procedure is deferred until section I, where precise measurement is discussed for joint prior distributions. The derivation of the estimator for μ_y in the case of precise measurement is presented here since the estimator is a function of two quantities Q_y and Q_x which occur in varying combinations throughout the remainder of this chapter.

2. <u>Sufficient statistics</u>

Lemma 7.1 below derives a set of sufficient statistics for the sampling scheme of this chapter when the posterior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ is desired.

Lemma 7.1. Let the sample data $(y_1, y_2, \ldots, y_{n+n_y}, x_1, x_2, \ldots, x_{n+n_x})$ be selected from a bivariate normal distribution according to the sampling specifications in section D of this chapter. Then, the statistics Q_x and Q_y are sufficient statistics for the Bayesian posterior analysis of $(\tilde{\mu}_x, \tilde{\mu}_y)$, where

$$Q_y = \frac{nn_x\left[\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x\right] + n^2\bar{y}_n + n_y\left[n + n_x(1-\rho^2)\right]\bar{y}_{n_y}}{\left[nn_x + n^2 + n_y\left\{n + n_x(1-\rho^2)\right\}\right]} \tag{7.5.1}$$

$$s_y^2 = \frac{\sigma_y^2}{M_y(1-R_{xy}^2)} \tag{7.5.2}$$

$$s_{xy}^2 = R_{xy}s_x s_y = s_{yx}^2 \tag{7.5.3}$$

$$M_y = n_y + \frac{n}{(1-\rho^2)} \tag{7.5.4}$$

$$R_{xy} = \frac{\rho n}{(1-\rho^2)\sqrt{M_x}\sqrt{M_y}} = R_{yx}. \tag{7.5.5}$$

Q_x, S_x^2, and M_x are obtained from Q_y, S_y^2, and M_y, respectively, by substituting x for y. The quantities S_y^2 and S_x^2 are shown later to be $V(\tilde{Q}_y)$ and $V(\tilde{Q}_x)$, respectively, and R_{xy} is shown to be the correlation between $\tilde{Q}_y$ and $\tilde{Q}_x$.

Proof: The conditional distribution of any one of the $(\tilde{x}_i, \tilde{y}_i)$, i = 1, 2, ..., n, is $f_1(x_i, y_i \mid \mu_x, \mu_y)$, where

$$f_1(x,y \mid \mu_x,\mu_y) = \left[2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}\right]^{-1} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\right\}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right] \tag{7.5.6}$$

The conditional distribution of any of the $\tilde{x}_i$, $i = n+1, \ldots, n+n_x$, is $f_2(x_i|\mu_x)$ where

$$f_2(x|\mu_x) = (\sigma_x\sqrt{2\pi})^{-1} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}. \tag{7.5.7}$$

The conditional distribution $f_3(y_i|\mu_y)$ of any of the $\tilde{y}_i$, $i=n+1, \ldots, n+n_y$, follows from equation (7.5.7) by replacing x with y.

The likelihood function L, conditional upon the parameters μ_x and μ_y, is

$$L = \prod_{i=1}^{n} f_1(x_i, y_i \mid \mu_x, \mu_y) \prod_{j=n+1}^{n+n_x} f_2(x_j \mid \mu_x) \prod_{k=n+1}^{n+n_y} f_3(y_k \mid \mu_y). \tag{7.5.8}$$

To obtain a set of sufficient statistics, it is necessary to consider only the kernel of the likelihood, i.e. only those terms of L which contain μ_x and/or μ_y.

Now, let

$$\begin{aligned} x_i - \mu_x &= x_i - \bar{x}_n + \bar{x}_n - \mu_x, \ i = 1, 2, \ldots, n. \\ x_j - \mu_x &= x_j - \bar{x}_{n_x} + \bar{x}_{n_x} - \mu_x, \ j = n+1, \ldots, n+n_x, \end{aligned} \tag{7.5.9}$$

and define $(y_i - \mu_y)$ in a similar manner. Substituting equation (7.5.9) into equation (7.5.8) to find the kernel of L yields, after some algebra,

$$L \propto \exp\left[\frac{-n(\bar{x}_n - \mu_x)^2}{2\sigma_x^2(1-\rho^2)} + \frac{\rho n(\bar{x}_n - \mu_x)(\bar{y}_n - \mu_y)}{\sigma_x \sigma_y (1-\rho^2)} - \frac{n(\bar{y}_n - \mu_y)^2}{2\sigma_y^2(1-\rho^2)} - \frac{n_x(\bar{x}_{n_x} - \mu_x)^2}{2\sigma_x^2} - \frac{n_y(\bar{y}_{n_y} - \mu_y)^2}{\sigma_y^2}\right] \tag{7.5.10}$$

In equation (7.5.10) the coefficient on μ_x^2 is $Co(\mu_x^2)$, where

$$-2Co(\mu_x^2) = \left[\frac{n}{\sigma_x^2(1-\rho^2)} + \frac{n_x}{\sigma_x^2}\right] = \frac{1}{S_x^2(1-R_{xy}^2)} \tag{7.5.11}$$

by using equations (7.5.1) through (7.5.5). The equation involving $Co(\mu_y^2)$ follows from equation (7.5.11) by substituting y for x. Also, after some algebra, it can be shown that

$$\mathrm{Co}(\mu_x\mu_y) = \frac{\rho n}{\sigma_x\sigma_y(1-\rho^2)} = \frac{R_{xy}}{S_xS_y(1-R_{xy}^2)} \tag{7.5.12}$$

and

$$\mathrm{Co}(\mu_x) = \frac{n\bar{x}_n}{(1-\rho^2)\sigma_x^2} - \frac{\rho n\bar{y}_n}{(1-\rho^2)\sigma_x\sigma_y} + \frac{n_x\bar{x}_{n_x}}{\sigma_x^2}$$

$$= \frac{1}{(1-R_{xy}^2)}\left[\frac{Q_x}{S_x^2} - \frac{R_{xy}Q_y}{S_xS_y}\right]. \tag{7.5.13}$$

The equation involving $\mathrm{Co}(\mu_y)$ can be obtained from equation (7.5.13) by substituting y for x. Using equations (7.5.10) through (7.5.13), the kernel of the likelihood function can be written as

$$L \propto \exp\left\{\frac{-1}{2(1-R_{xy}^2)}\right\}\left[\left(\frac{\mu_x-Q_x}{S_x}\right)^2 + \left(\frac{\mu_y-Q_y}{S_y}\right)^2 - 2R_{xy}\left(\frac{\mu_x-Q_x}{S_x}\right)\left(\frac{\mu_y-Q_y}{S_y}\right)\right] \tag{7.5.14}$$

since the coefficients on μ_x^2, μ_y^2, $\mu_x\mu_y$, μ_x, and μ_y are the same in equations (7.5.10) and (7.5.14). By the factorability criterion for sufficiency in the Bayesian sense, equation (7.5.14) shows that Q_x and Q_y are sufficient statistics for the posterior analysis of $\tilde{\mu}_x$ and $\tilde{\mu}_y$. This completes the proof to Lemma 7.1.

By Lemma 3.1 of Chapter III, the joint distribution of $\tilde{Q}_x$ and $\tilde{Q}_y$, given μ_x and μ_y, is bivariate normal. Also, from equation (7.5.1) the conditional expectation of $\tilde{Q}_x$ and $\tilde{Q}_y$ is easily found to be

$$E\begin{bmatrix}\tilde{Q}_x & \mu_x \\ \tilde{Q}_y & \mu_y\end{bmatrix} = \begin{bmatrix}\mu_x \\ \mu_y\end{bmatrix}. \tag{7.5.15}$$

Using equations (7.5.1) through (7.5.5) and (3.3.3), a little algebra yields the covariance matrix of $(\tilde{Q}_x, \tilde{Q}_y)$ as

$$V\begin{bmatrix}\tilde{Q}_x & \mu_x \\ \tilde{Q}_y & \mu_y\end{bmatrix} = \begin{bmatrix}S_x^2 & S_{xy}^2 \\ S_{xy}^2 & S_y^2\end{bmatrix}. \tag{7.5.16}$$

Since Q_x and Q_y are sufficient statistics for the posterior analysis of $\tilde{\mu}_x$ and $\tilde{\mu}_y$, then the entire Bayesian analysis can be done by using only the distribution of $\tilde{Q}_x$ and $\tilde{Q}_y$, given μ_x and μ_y, rather than the conditional distribution of all the sample data. Sufficiency implies that Q_x nnd Q_y must contain all of the information from the sample about the parameters μ_x and μ_y.

Note that Q_y in equation (7.5.1) is a weighted average of three common estimators of μ_y: the mean $\bar{y}_n$, and mean $\bar{y}_{n_y}$, and the regression estimator $\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x$ where $\bar{x}_{n_x}$ replaces the usual population mean μ_x, or in cases of double sampling, replaces an estimate of μ_x based on a larger sample than the sample that yields $\bar{x}_n$ and $\bar{y}_n$.

3. <u>Posterior distribution of $\tilde{\mu}_y$ and $\tilde{\mu}_x$</u>

Assuming that the prior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ can be regarded as constant, then the posterior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ is proportional to the kernel of the likelihood function given in equation (7.5.14), i.e.

$$h(\mu_x, \mu_y \mid Q_x, Q_y) \propto \exp\left\{-\frac{1}{2(1-R_{xy}^2)}\right\} \left[\begin{array}{l} \left(\frac{\mu_x - Q_x}{S_x}\right)^2 + \left(\frac{\mu_y - Q_y}{S_y}\right)^2 \\ - 2R_{xy}\left(\frac{\mu_x - Q_x}{S_x}\right)\left(\frac{\mu_y - Q_y}{S_y}\right) \end{array}\right]. \tag{7.5.17}$$

By inspection, it can be seen that the posterior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ is thus bivariate normal with mean

$$E\left[\begin{array}{c|c} \tilde{\mu}_x & Q_x \\ \tilde{\mu}_y & Q_y \end{array}\right] = \begin{bmatrix} Q_x \\ Q_y \end{bmatrix} \tag{7.5.18}$$

and covariance matrix

$$V\left[\begin{array}{c|c} \tilde{\mu}_x & Q_x \\ \tilde{\mu}_y & Q_y \end{array}\right] = \begin{bmatrix} S_x^2 & S_{xy}^2 \\ S_{xy}^2 & S_y^2 \end{bmatrix}. \tag{7.5.19}$$

Hence, the marginal posterior distribution of $\tilde{\mu}_y$ is normal with mean Q_y and variance S_y^2.

4. <u>Loss function and estimator of μ_y</u>

Letting $\hat{\mu}_y$ be the estimator of μ_y, the loss function which is considered throughout this chapter is

$$\text{Loss}(\mu_y, \hat{\mu}_y) = b(\mu_y - \hat{\mu}_y)^2 + nC_{xy} + n_yC_y + n_xC_x + C^* \tag{7.5.20}$$

where $b > 0$ is a constant, C^* is the overhead cost, and C_{xy}, C_y, and C_x are the sampling costs as defined in section C of Chapter III. The estimator $\hat{\mu}_y = E\left[\tilde{\mu}_y \mid Q_x, Q_y\right]$ minimizes $E\left[\text{Loss}(\tilde{\mu}_y, \hat{\mu}_y) \mid Q_x, Q_y\right]$ where expectation is taken with respect to the posterior distribution of $\tilde{\mu}_y$. For the rest of this chapter, the posterior mean of $\tilde{\mu}_y$ will be taken as the estimator of $\tilde{\mu}_y$.

Thus, the estimator $\tilde{\mu}_y$ for this section is

$$\tilde{\mu}_y = E(\tilde{\mu}_y \mid Q_y) = Q_y \tag{7.5.21}$$

where Q_y is given in equation (7.5.1). It has been noted that Q_y is a weighted average of three estimators of μ_y--two means and one regression estimator. It can be shown that the weights on the three different estimators in Q_y minimize the variance of the quantity

$$w_1\left[\bar{y}_n + \beta(\bar{x}_{n_x} - \bar{x}_n)\right] + w_2\bar{y}_n + w_3\bar{y}_{n_y} \quad ,$$

where $(w_1 + w_2 + w_3) = 1$. An investigation of Q_y when some of the sample sizes are zero also gives some intuitive justification for the weights on the three components of Q_y. For example, if $n_x = 0$, then

$$Q_y = \left[n\bar{y}_n + n_y\bar{y}_{n_y}\right] / (n+n_y), \; n_x = 0. \tag{7.5.22}$$

Thus, Q_y no longer contains a regression estimator of μ_y and therefore does not incorporate the information $\bar{x}_n$ since there is no additional information on $\tilde{x}$ with which to compare $\bar{x}_n$.

If $n_y = 0$, then

$$Q_y = \frac{n\bar{y}_n + n_x\left[\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x\right]}{(n+n_x)}, \; n_y = 0 \tag{7.5.23}$$

and all available sample data is incorporated into the estimator of μ_y.

In this case Q_y is a weighted average of $\bar{y}_n$ and a regression estimator. If $n_y = n_x = 0$, then

$$Q_y = \bar{y}_n, \quad n_x = n_y = 0, \tag{7.5.24}$$

and the information $\bar{x}_n$ is not incorporated into Q_y since a regression estimator is not possible. If $n = 0$, then

$$Q_y = \bar{y}_{n_y}, \; n = 0, \tag{7.5.25}$$

and Q_y is not a function of $\bar{x}_{n_x}$ because a regression estimator is not possible. Thus, in equations (7.5.24) and (7.5.25), note that none of the available sample data on $\tilde{x}$ is incorporated into Q_y.

Hence, the observations on $\tilde{x}$ are used in the estimator of μ_y only when $n > 0$ and $n_x > 0$, since the regression term is of the form $(\bar{x}_{n_x} - \bar{x}_n)$. Note also in equation (7.5.1) that if $n_x < n$, the regression estimator is weighted less than the mean $\bar{y}_n$.

Substituting $\hat{\mu}_y = E(\tilde{\mu}_y | Q_y)$ into the loss function in equation (7.5.20) yields the posterior loss as

$$E\left[\text{Loss}(\tilde{\mu}_y, \hat{\mu}_y \mid Q_y)\right] = bS_y^2 + nC_{xy} + n_yC_y + n_xC_x + C^*. \tag{7.5.28}$$

Note that the posterior loss does not depend upon the sample data.

Assume now that a fixed budget Co is available for sampling, so that

$$C_o = nC_{xy} + n_yC_y + n_xC_x. \tag{7.5.29}$$

Then it is desireble to minimize the posterior loss in equation (7.5.28) by selecting sample sizes $n \geq 0$, $n_y \geq 0$, and $n_x \geq 0$ such that equation

(7.5.29) is satisfied. This problem is equivalent to the problem of maximizing σ_y^2 / S_y^2 subject to the constraints

$$n \geq 0$$

$$n_x \geq 0 \qquad (7.5.30)$$

$$nC_{xy} + n_x C_x \leq C_o,$$

where

$$\frac{\sigma_y^2}{S_y^2} = \frac{C_o - nC_{xy} - n_x C_x}{C_y} \div \frac{n(n+n_x)}{\left[n+n_x(1-\rho^2)\right]} . \qquad (7.5.31)$$

The usual calculus techniques do not offer a solution. Also, it can be shown that the function σ_y^2/S_y^2 is not a convex function, and thus convex programming will not offer a solution to the problem. It appears that the best approach to finding an optimum allocation is to consider the triangular region in the (n, n_x) plane as indicated below in Figure 7.2 and evaluate σ_y^2/S_y^2 in some systematic manner until an optimum solution is obtained.

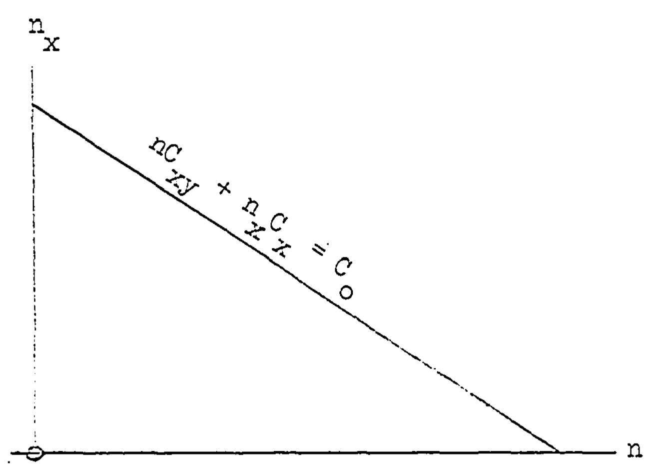

Figure 7.2. The admissable region for the optimum sample allocation

F. Bivariate Normal Prior on $\tilde{\mu}_y$ and $\tilde{\mu}_x$

1. Prior distribution on $\tilde{\mu}_y$ and $\tilde{\mu}_x$

In this section the prior distribution on $(\tilde{\mu}_x, \tilde{\mu}_y)$ is assumed to be bivariate normal with mean

$$E\begin{bmatrix}\tilde{\mu}_x \\ \tilde{\mu}_y\end{bmatrix} = \begin{bmatrix} b_x \\ b_y \end{bmatrix} \qquad (7.6.1)$$

and covariance matrix

$$V\begin{bmatrix}\tilde{\mu}_x \\ \tilde{\mu}_y\end{bmatrix} = \begin{bmatrix} \alpha_x^2 & \gamma\alpha_x\alpha_y \\ \gamma\alpha_x\alpha_y & \alpha_y^2 \end{bmatrix}. \qquad (7.6.2)$$

This is also the natural conjugate prior distribution of the bivariate normal distribution when the covariance matrix of $(\tilde{x}, \tilde{y})$ is known. The estimator of μ_y derived under this prior distribution is the most general estimator in this chapter, and most other estimators which are discussed later on in this chapter, including the estimators which pool means, can be considered as special cases of this general estimator.

2. Posterior distribution of $\tilde{\mu}_y$ and $\tilde{\mu}_x$

Since interest in this discussion is on the posterior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ as in section E, the sufficient statistics Q_x and Q_y of Lemma 7.1 can be used to summarize the sample data. Thus, it is necessary to find the joint posterior distribution $h(\mu_x, \mu_y \mid Q_x, Q_y)$ when the data distribution $f(Q_x, Q_y \mid \mu_x, \mu_y)$ and the prior distribution $g(\mu_x, \mu_y)$ are both bivariate normal.

Now, theorem 12.1.4 of Raiffa and Schlaifer (1961, pp. 311-312) states the following: if the vector $\underline{x}$, given the unknown mean vector $\underline{\mu}$ and known covariance matrix Σ, is multivariate normal, and if the prior distribution on the vector $\underline{\mu}$ is multivariate normal with parameters

$$E\left[\tilde{\underline{\mu}}\right] = \underline{\mu}_p \qquad V\left[\tilde{\underline{\mu}}\right] = \Sigma_p, \tag{7.6.3}$$

then the posterior distribution of $\tilde{\underline{\mu}}$ is multivariate normal with parameters

$$E\left[\tilde{\underline{\mu}} \mid \underline{x}\right] = \left[\Sigma_p^{-1} + \Sigma^{-1}\right]^{-1} \left[\Sigma_p^{-1}\, \underline{\mu}_p + \Sigma^{-1}\underline{x}\right]$$

$$V\left[\tilde{\underline{\mu}} \mid \underline{x}\right] = \left[\Sigma_p^{-1} + \Sigma^{-1}\right]^{-1} . \tag{7.6.4}$$

In the application of this theorem to finding $h(\mu_x, \mu_y \mid Q_x, Q_y)$, let

$$\underline{x} = \begin{bmatrix} Q_x \\ Q_y \end{bmatrix} \tag{7.6.5}$$

$$\underline{\mu}_p = \begin{bmatrix} b_x \\ b_y \end{bmatrix} \tag{7.6.6}$$

$$\Sigma = \begin{bmatrix} s_x^2 & s_{xy}^2 \\ s_{xy}^2 & s_y^2 \end{bmatrix} \tag{7.6.7}$$

$$\Sigma_p = \begin{bmatrix} \alpha_x^2 & \gamma\alpha_x\alpha_y \\ \gamma\alpha_x\alpha_y & \alpha_y^2 \end{bmatrix} \tag{7.6.8}$$

Then,

$$\Sigma_p^{-1} = \frac{1}{\alpha_x^2\alpha_y^2(1-\gamma^2)} \begin{bmatrix} \alpha_y^2 & -\gamma\alpha_x\alpha_y \\ -\gamma\alpha_x\alpha_y & \alpha_x^2 \end{bmatrix} \tag{7.6.9}$$

and

$$\Sigma^{-1} = \frac{1}{s_x^2 s_y^2(1-R_{xy}^2)} \begin{bmatrix} s_y^2 & -s_{xy}^2 \\ -s_{xy}^2 & s_x^2 \end{bmatrix}. \tag{7.6.10}$$

Thus, the posterior covariance matrix of μ_x and μ_y is

$$V\begin{bmatrix} \tilde{\mu}_x & Q_x \\ \tilde{\mu}_y & Q_y \end{bmatrix} = \frac{1}{D}\begin{bmatrix} \dfrac{1}{\alpha_y^2(1-\gamma^2)} + \dfrac{1}{s_y^2(1-R_{xy}^2)} & \dfrac{\gamma}{\alpha_x\alpha_y(1-\gamma^2)} + \dfrac{R_{xy}}{s_x s_y(1-R_{xy}^2)} \\ \dfrac{\gamma}{\alpha_x\alpha_y(1-\gamma^2)} + \dfrac{R_{xy}}{s_x s_y(1-R_{xy}^2)} & \dfrac{1}{\alpha_x^2(1-\gamma^2)} + \dfrac{1}{s_x^2(1-R_{xy}^2)} \end{bmatrix} \tag{7.6.11}$$

where D, the determinant of $\left[\Sigma_p^{-1} + \Sigma^{-1}\right]$, is

$$D = \left[\frac{1}{\alpha_x^2(1-\gamma^2)} + \frac{1}{s_x^2(1-R_{xy}^2)}\right]\left[\frac{1}{\alpha_y^2(1-\gamma^2)} + \frac{1}{s_y^2(1-R_{xy}^2)}\right] - \left[\frac{\gamma}{\alpha_x\alpha_y(1-\gamma^2)} + \frac{R_{xy}}{s_x s_y(1-R_{xy}^2)}\right]^2 . \tag{7.6.12}$$

After performing the matrix multiplication to obtain $E(\tilde{\underline{\mu}} \mid \underline{x})$, some additional algebra yields the posterior means as

$$E(\tilde{\mu}_y \mid Q_x, Q_y) = \frac{1}{D}\left[\frac{b_y}{\alpha_x\alpha_y(1-\gamma^2)}\left\{\frac{1}{\alpha_x\alpha_y} - \frac{\gamma R_{xy}}{S_xS_y(1-R_{xy}^2)}\right\} + \frac{Q_y}{S_xS_y(1-R_{xy}^2)}\left\{\frac{1}{S_xS_y} - \frac{\gamma R_{xy}}{\alpha_x\alpha_y(1-\gamma^2)}\right\} + \frac{\left[Q_y + R_{xy}S_y(b_x - Q_x)/S_x\right]}{S_y^2\alpha_x^2(1-\gamma^2)(1-R_{xy}^2)} + \frac{\left[b_y + \gamma\alpha_y(Q_x - b_x)/\alpha_x\right]}{S_x^2\alpha_y^2(1-\gamma^2)(1-R_{xy}^2)}\right] \quad (7.6.13)$$

The posterior mean $E(\tilde{\mu}_x \mid Q_x, Q_y)$ can be obtained from equation (7.6.13) by substituting x for y.

Equation (7.6.13) above gives the most general posterior mean of $\tilde{\mu}_y$ in this chapter, and it is worth while to investigate the components of the estimator to see how the prior and sample information combine. First, note that the posterior mean of $\tilde{\mu}_y$ is a weighted average of four components: the prior mean of $\tilde{\mu}_y$, i.e. b_y; an unbiased sample estimator of μ_y, i.e. Q_y; a regression of Q_y upon Q_x, where b_x is in the position usually occupied by μ_x; and a regression of b_y upon b_x, where Q_x now is in the position usually occupied by μ_x. Some of the weights on the components can be interpreted intuitively. Consider the third component of the posterior mean, i.e. the regression of Q_y upon Q_x. The weight on this regression term increases as $S_y^2(1-R_{xy}^2)$ decreases. Note that $S_y^2(1-R_{xy}^2)$ is the usual conditional variance of $\tilde{Q}_y$, given $\tilde{Q}_x = Q_x$. Also, the weight on this regression term increases as α_x^2 decreases. Thus, the more certain it is that b_x is a suitable replacement for μ_x in the regression term, the larger the weight on the regression term. A similar interpretation holds for the fourth component of the posterior mean, the regression of b_y upon b_x. As S_x^2 increases, the weight on the fourth component decreases since Q_x is then not a reliable substitute for μ_x in the regression term. Also, as $\alpha_y^2(1-\gamma^2)$ increases, the weight on the fourth component decreases since b_y is then not a reliable estimator of μ_y. In the first component, the weight on b_y increases as α_y^2 decreases. In the second component of the posterior mean, the weight on Q_y increases as S_y^2 decreases.

Recall from equation (7.5.1) that both Q_y and Q_x incorporate the observations on $\tilde{x}$ when $\rho \neq 0$. However, in equation (7.6.13), Q_x is used only in the regression component of the posterior mean, whereas Q_y is used both by itself and also in the regression component. A similar situation holds for b_y and b_x in equation (7.6.13).

3. <u>Loss function and estimator of μ_y</u>

Since the posterior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$ is bivariate normal, then the marginal posterior distribution of $\tilde{\mu}_y$ is normal with mean and variance as given in equations (7.6.13), (7.6.12), and (7.6.11). Assuming the quadratic loss function as given in equation (7.5.20), the estimator $\hat{\mu}_y$ is the posterior mean $E(\tilde{\mu}_y \mid Q_x, Q_y)$ as given in equation (7.6.13).

It is easily shown that if $\alpha_x^2 \longrightarrow \infty$ and $\alpha_y^2 \longrightarrow \infty$, then $E(\tilde{\mu}_y \mid Q_x, Q_y)$ in equation (7.6.13) approaches Q_y. Letting α_x^2 and α_y^2 approach infinity flattens out the bivariate normal prior density on $(\tilde{\mu}_x, \tilde{\mu}_y)$, and the estimator $\hat{\mu}_y$ in equation (7.6.13) approaches the estimator Q_y obtained for the case when the prior distribution on $(\tilde{\mu}_x, \tilde{\mu}_y)$ was assumed to be constant. Thus, the estimator Q_y in section E can be justified by considering vague or nonexistent prior information to be represented by a bivariate normal prior distribution on $(\tilde{\mu}_x, \tilde{\mu}_y)$ with extremely large variances α_x^2 and α_y^2.

With $\hat{\mu}_y$ as in equation (7.6.13), the expected conditional loss from equation (7.5.20) is

$$E\left[\text{Loss}(\tilde{\mu}_y, \hat{\mu}_y \mid Q_x, Q_y)\right] = bV(\tilde{\mu}_y \mid Q_x, Q_y) + nC_{xy} + n_y C_y + n_x C_x + C^* \tag{7.6.14}$$

where $V(\tilde{\mu}_y \mid Q_x, Q_y)$ is given in equations (7.6.11) and (7.6.12). Note that the expected conditonal loss is independent of the sample results.

An attempt to obtain a sample allocation of n, n_y, and n_x to minimize the posterior loss in equation (7.6.14) subject to a fixed sampling budget runs into the same difficulties discussed in the latter part of section E. The makeshift solution offered there is also appropriate here with obvious modifications to include now the parameters of the prior distribution on $(\tilde{\mu}_y, \tilde{\mu}_x)$.

G. Special Case Where $\tilde{x}$ and $\tilde{y}$ Are Independent

1. Sampling scheme

Consider now the important special case where $\tilde{x}$ and $\tilde{y}$ are independent. $(\tilde{\mu}_x, \tilde{\mu}_y)$ are still assumed to have a bivariate normal prior distribution with parameters as given in equations (7.6.1) and (7.6.2).

The general results of section F of this chapter can be specialized to give the estimator of μ_y for this special case. Since $\rho = 0$, then there is no matched, bivariate sample. For simplicity, then, let $n = 0$ and consider the data to consist of n_x observations on $\tilde{x}$ and n_y observations on $\tilde{y}$. Thus, all the formulas of section F apply here when n and ρ are equated to zero.

2. Estimator of μ_y

Since now $\rho = 0$ and $n = 0$, then equations (7.5.1) through (7.5.5) imply that

$$\begin{aligned} Q_y &= \bar{y}_{n_y} \\ Q_x &= \bar{x}_{n_x} \\ R_{xy} &= 0 = S_{xy}^2 \qquad (\rho = n = 0) \qquad (7.7.1) \\ S_y^2 &= \sigma_y^2/n_y \\ S_x^2 &= \sigma_x^2/n_x \end{aligned}$$

From section F, the posterior distribution of $\tilde{\mu}_y$ is normal with mean

$$E(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \frac{1}{D}\left[\begin{array}{l} \dfrac{b_y}{\alpha_x^2\alpha_y^2(1-\gamma^2)} \\ + \dfrac{\bar{y}_{n_y} n_y}{\sigma_y^2}\left\{\dfrac{n_x}{\sigma_x^2} + \dfrac{1}{\alpha_x^2(1-\gamma^2)}\right\} \\ + \dfrac{n_x\left[b_y + \gamma\alpha_y(\bar{x}_{n_x} - b_x)/\alpha_x\right]}{\sigma_x^2\alpha_y^2(1-\gamma^2)} \end{array}\right] \tag{7.7.2}$$

and variance

$$V(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \frac{1}{D}\left\{\frac{1}{\alpha_x^2(1-\gamma^2)} + \frac{n_x}{\sigma_x^2}\right\} \tag{7.7.3}$$

where

$$D = \frac{1}{\alpha_x^2\alpha_y^2(1-\gamma^2)} + \frac{n_y}{\sigma_y^2}\left\{\frac{1}{\alpha_x^2(1-\gamma^2)} + \frac{n_x}{\sigma_x^2}\right\} + \frac{n_x}{\sigma_x^2\alpha_y^2(1-\gamma^2)} \quad . \tag{7.7.4}$$

From equation (7.5.20) the loss function when $\tilde{x}$ and $\tilde{y}$ are independent is

$$\mathrm{Loss}(\mu_y, \hat{\mu}_y) = b(\mu_y - \hat{\mu}_y)^2 + n_y C_y + n_x C_x + C^*, \tag{7.7.5}$$

where the estimator $\hat{\mu}_y$ for this special case is

$$\hat{\mu}_y = E(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) \tag{7.7.6}$$

as given in equation (7.7.2).

$\hat{\mu}_y$ is a weighted average of three estimators: the prior mean b_y, the sample mean $\bar{y}_{n_y}$, and the regression of b_y upon b_x. The observations on $\tilde{x}$ are used only in the regression component of the estimator. Recall from equation (7.6.13) that the general estimator is a weighted average of Q_y, b_y, the regression of Q_y upon Q_x, and the regression of b_y upon b_x. Note, in equation (7.7.2), where now $\tilde{x}$ and $\tilde{y}$ are independent, that the regression of $Q_y = \bar{y}_{n_y}$ upon $Q_x = \bar{x}_{n_x}$ no longer appears. Obviously, it is not reasonable to correct $\bar{y}_{n_y}$ by means of a regression upon $\bar{x}_{n_x}$ when $\tilde{\bar{y}}_{n_y}$ and $\tilde{\bar{x}}_{n_x}$ are independent. Further, a moderate amount of algebra shows that the weights on the three components of μ_y in equation (7.7.2) minimize the "variance" of μ_y. The "variance" of μ_y is obtained by considering b_y and b_x as random variables with variances α_y^2 and α_x^2, respectively, and covariance $\gamma\alpha_y\alpha_x$. The variances of $\bar{x}_{n_x}$ and $\bar{y}_{n_x}$ are taken with respect to the conditional distribution of $(\tilde{x}, \tilde{y})$, given (μ_x, μ_y), and (b_y, b_x) is assumed to be independent of $(\tilde{\bar{x}}_{n_x}, \tilde{\bar{y}}_{n_y})$. This "variance" of μ_y as explained above is not a typical Bayesian concept, but such an analysis helps in comparing the Bayesian estimator to the usual classical estimators.

Some intuitive interpretation of the weights on the three components in equation (7.7.2) can be made upon inspection. As α_y^2 increases, the weight on b_y decreases, expressing the uncertainty that b_y is a reliable estimator of μ_y. Second, as σ_y^2/n_y increases, the weight on $\bar{y}_{n_y}$ decreases. Third, the weight on the regression component decreases as σ_x^2/n_x increases and $\alpha_y^2(1-\gamma^2)$ increases. If σ_x^2/n_x is large, then $\bar{x}_{n_x}$ is not a reliable replacement for μ_x in the regression term. Likewise, if $\alpha_y^2(1-\gamma^2)$ is large, either b_y is not a reliable estimator of μ_y or the correlation γ between $\tilde{b}_y$ and $\tilde{b}_x$ is not high enough to warrant a significant contribution from the regression of b_y upon b_x. Note, finally, that if $\gamma = 0$, i.e. the prior distributions on $\tilde{\mu}_x$ and $\tilde{\mu}_y$ are independent, then μ_y is simply a weighted average of b_y and $\bar{y}_{n_y}$, and no use is made of the observations on $\tilde{x}$ in the estimation of μ_y.

From equation (7.7.5), the posterior loss is

$$E\left[\text{Loss}(\tilde{\mu}_y, \hat{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})\right] = bV(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) + n_y C_y + n_x C_x + C^* \tag{7.7.7}$$

where $V(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})$ is given in equation (7.7.3). Note that the posterior loss is independent of the sample data $\bar{x}_{n_x}$ and $\bar{y}_{n_y}$.

3. Optimum sample allocation

Let a fixed budget C_o be available for sampling, so that

$$C_o = n_x C_x + n_y C_y. \tag{7.7.8}$$

It is desired to select n_x and n_y such that the posterior loss in equation (7.7.7) is minimized. Define now

$$A = \frac{1}{\alpha_x^2(1-\gamma^2)} + \frac{C_o - n_y C_y}{C_x \sigma_x^2} = DV(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})$$

$$B = \frac{1}{\alpha_y^2(1-\gamma^2)} + \frac{n_y}{\sigma_y^2} = DV(\tilde{\mu}_x \mid \bar{x}_{n_x}, \bar{y}_{n_y}) \tag{7.7.9}$$

$$C = \gamma\left[\alpha_x\alpha_y(1-\gamma^2)\right]^{-1} = D\ \mathrm{cov}(\tilde{\mu}_x, \tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})$$

where D is given in equation (7.7.4) and, in addition, satisfies

$$D = AB - C^2. \tag{7.7.10}$$

Then, from equation (7.7.7)

$$E\left[\mathrm{Loss}\ (\tilde{\mu}_y, \hat{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})\right] = bA(AB-C^2)^{-1} + C_o + C^*. \tag{7.7.11}$$

The right hand side of equation (7.7.11) is a function of n_y, and the two critical values of n_y are found to be

$$n_y = \frac{C_o}{C_y} + \frac{C_x\sigma_x^2}{C_y\alpha_x^2(1-\gamma^2)} \pm \frac{\sigma_y\sigma_x\gamma\sqrt{C_x}}{\alpha_x\alpha_y(1-\gamma^2)\sqrt{C_y}}. \tag{7.7.12}$$

Clearly, one of the above critical values is always larger than C_o/C_y and is thus not an admissable solution. The remaining solution n_{yc} may be an admissable solution, where

$$n'_{yc} = \frac{C_o}{C_y} + \frac{C_x\sigma_x^2}{C_y\alpha_x^2(1-\gamma^2)} - \frac{|\gamma|\ \sigma_x\sigma_y\sqrt{C_x}}{\alpha_x\alpha_y(1-\gamma^2)\sqrt{C_y}}. \tag{7.7.13}$$

The second derivative of the posterior loss with respect to n_y is

$$\frac{d^2E}{dn_y^2}\left[\mathrm{Loss}\ (\tilde{\mu}_y, \hat{\mu}_y) \mid \bar{x}_{n_x}, \bar{y}_{n_y})\right] = 2(AB-C^2)^{-3}\left\{\frac{A^3}{\sigma_y^4} - \frac{2AC^2C_y}{\sigma_x^2\sigma_y^2C_x} + \frac{BC^2C_y^2}{\sigma_x^4C_x^2}\right\}. \tag{7.7.14}$$

Assume now that the critical value n_{yc} is admissable, i.e.

$$0 \le n_{yc} \le C_o/C_y. \tag{7.7.15}$$

Then, for $n_y = n_{yc}$ satisfying inequality (7.7.15), $A > 0$ and $B > 0$ from equation (7.7.9). Hence $D > 0$, since A/D and B/D are diagonal elements of the covariance matrix of the posterior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$. Thus

$$B > C^2/A. \tag{7.7.16}$$

Now, for $n_y = n_{yc}$ satisfying inequality (7.7.15),

$$\frac{A^3}{\sigma_y^4} - \frac{2AC^2c_y}{\sigma_x^2\sigma_y^2c_x} + \frac{BC^2c_y^2}{\sigma_x^4c_x^2}$$

$$> \frac{A^3}{\sigma_y^4} - \frac{2AC^2c_y}{\sigma_x^2\sigma_y^2c_x} + \frac{C^4c_y^2}{A\sigma_x^4c_x^2}$$

$$= \frac{1}{A}\left\{\frac{A^2}{\sigma_y^2} - \frac{C^2c_y}{\sigma_x^2c_x}\right\}^2 = 0. \tag{7.7.17}$$

Thus, for $0 \leq n_{yc} \leq c_o/c_y$,

$$\frac{d^2}{dn_y^2} V(\tilde{\mu}_y \mid Q_x, Q_y) > 0 \tag{7.7.18}$$

and thus $n_y = n_{yc}$ yields a minimum value for the posterior loss.

An analysis of the posterior loss in equation (7.7.11) shows that it can assume three different shapes for various values of the known parameters. Figure 7.3 illustrates the three possible forms of the posterior loss in the relevant range $\left[0, c_o/c_y\right]$.

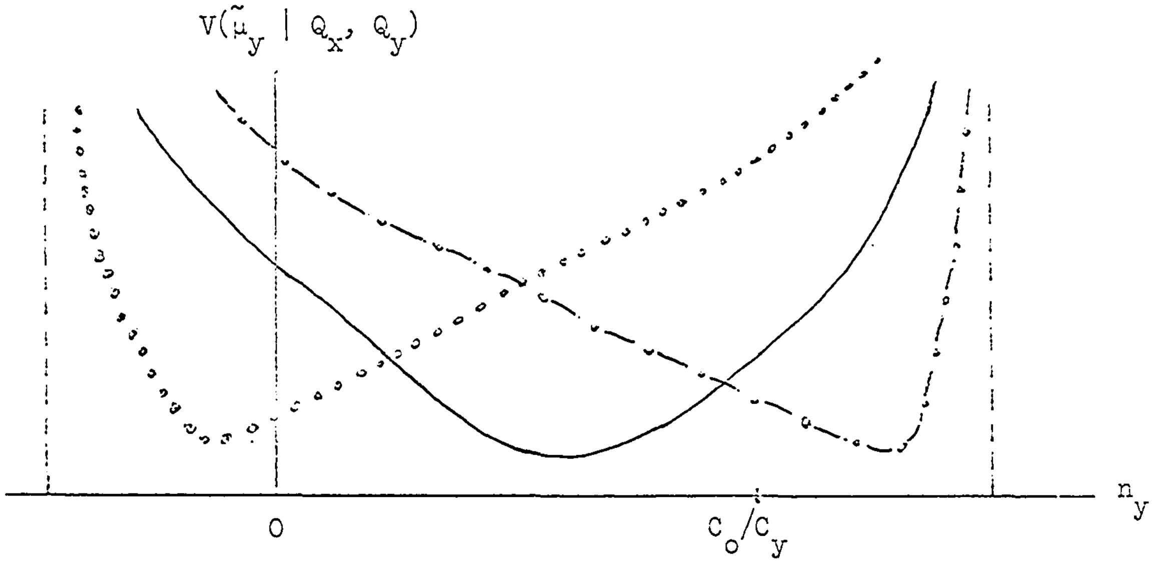

Figure 7.3. The three possible graphs for $E\left[\text{Loss}\ (\tilde{\mu}_y, \hat{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y})\right]$

The optimum value for n_y is thus given as follows. From equation (7.7.13) evaluate the critical value n_{yc}. Then,

$$n_y = n_{yc} \text{ if } 0 \leq n_{yc} \leq C_o/C_y$$

$$n_y = 0 \text{ if } n_{yc} < 0 \tag{7.7.19}$$

$$n_y = C_o/C_y \text{ if } n_{yc} > C_o/C_y\,.$$

With n_y as in equation (7.7.19), and

$$n_x = (C_o - C_y n_y)/C_x, \tag{7.7.20}$$

then the posterior loss will be minimized for a fixed sampling cost C_o.

H. Special Case Where $\tilde{\mu}_y$ and $\tilde{\Delta}$ are Independently Normally Distributed

1. Prior distribution on $\tilde{\mu}_y$ and $\tilde{\Delta}$

In the remaining sections of this chapter the prior distribution will be on $(\tilde{\mu}_y, \tilde{\Delta})$ rather than on $(\tilde{\mu}_y, \tilde{\mu}_x)$ as in sections F, G, and H, where $\Delta = \mu_y - \mu_x$. In many cases it is simpler and more meaningful to think of a prior distribution on $\tilde{\mu}_y$ and then of an independent prior distribution on the difference $\tilde{\Delta}$. For example, let $\tilde{y}$ be the weight of brand Y cigarette being produced by a manufacturing process, with a similar definition for $\tilde{x}$. Then, if it is desired to estimate the mean weight of brand Y, i.e. μ_y, it may be entirely reasonable to have a prior distribution on $\tilde{\mu}_y$ and then to have an independent prior distribution on the difference $\tilde{\Delta}$. Since the pooling of observations on $\tilde{x}$ and $\tilde{y}$ is being admitted as a possibility for the estimation of μ_y, then the prior distribution on $\tilde{\Delta}$ must reflect the assumption that μ_y and μ_x may be equal. Thus, let $E(\tilde{\Delta}) = 0$, and then $V(\tilde{\Delta}) = a^2$ reflects the prior uncertainty that μ_y and μ_x are equal.

In this section, the prior distribution on $\tilde{\mu}_y$ is assumed to be $N(by, \alpha_y^2)$, and the prior distribution on $\tilde{\Delta}$ is assumed to be $N(0, a^2)$. In addition, $\tilde{\mu}_y$ and $\tilde{\Delta}$ are assumed to be independently distributed.

The results of section F, where a bivariate normal prior on $(\tilde{\mu}_y, \tilde{\mu}_x)$ was assumed, can be used here by the appropriate selection of the elements of the mean vector and covariance matrix in equations (7.6.1) and (7.6.2).

For use in this section, let

$$b_x = b_y$$

$$\gamma = \alpha_y/\alpha_x \tag{7.8.1}$$

$$\alpha_x^2 = a^2 + \alpha_y^2.$$

With these specifications, it can be easily shown that

$$E(\tilde{\Delta}) = E(\tilde{\mu}_y - \tilde{\mu}_x) = 0$$

$$E(\tilde{\mu}_y) = b_y$$

$$\mathrm{Cov}(\tilde{\Delta}, \tilde{\mu}_y) = 0 \tag{7.8.2}$$

$$V(\tilde{\mu}_y) = \alpha_y^2$$

$$V(\tilde{\Delta}) = a^2,$$

and thus the desired properties of $\tilde{\mu}_y$ and $\tilde{\Delta}$ are obtained. Conversely, it is easily shown that equation (7.8.2) implies equation (7.8.1) if α_x^2, α_y^2, and $\gamma\alpha_x\alpha_y$ are defined to be the prior variances and prior covariance of $\tilde{\mu}_x$ and $\tilde{\mu}_y$, and if $\tilde{\Delta}$ and $\tilde{\mu}_y$ are assumed to be normally distributed.

2. Posterior distribution of $\tilde{\mu}_y$

If now the prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ are independently normally distributed with parameters given in equation (7.8.2), then the posterior distribution of $\tilde{\mu}_x$ and $\tilde{\mu}_y$ is bivariate normal with parameters given by equations (7.6.11) through (7.6.13) with the substitutions indicated in equation (7.8.1). Thus, the posterior covariance matrix is

$$V\left[\begin{array}{c|c} \tilde{\mu}_x & Q_x \\ \tilde{\mu}_y & Q_y \end{array}\right] = \frac{1}{D}\begin{bmatrix} \dfrac{1}{a^2}+\dfrac{1}{\alpha_y^2}+\dfrac{1}{S_y^2(1-R_{xy}^2)} & \dfrac{1}{a^2}+\dfrac{R_{xy}}{S_xS_y(1-R_{xy}^2)} \\ \dfrac{1}{a^2}+\dfrac{R_{xy}}{S_xS_y(1-R_{xy}^2)} & \dfrac{1}{a^2}+\dfrac{1}{S_x^2(1-R_{xy}^2)} \end{bmatrix} \tag{7.8.3}$$

where D, obtained from equation (7.6.12), is

$$D = \left\{\frac{1}{a^2}+\frac{1}{S_x^2(1-R_{xy}^2)}\right\}\left\{\frac{1}{a^2}+\frac{1}{\alpha_y^2}+\frac{1}{S_y^2(1-R_{xy}^2)}\right\} - \left\{\frac{1}{a^2}+\frac{R_{xy}}{S_xS_y(1-R_{xy}^2)}\right\}^2. \tag{7.8.4}$$

The posterior mean of $\tilde{\mu}_y$ is

$$E(\tilde{\mu}_y \mid Q_x, Q_y) = \frac{1}{D}\left\{\begin{aligned} & b_y\left[\frac{1}{\alpha_y^2a^2}-\frac{R_{xy}}{a^2S_xS_y(1-R_{xy}^2)}+\frac{1}{S_x^2\alpha_y^2(1-R_{xy}^2)}\right] \\ & +\frac{Q_y}{S_xS_y(1-R_{xy}^2)}\left[\frac{1}{S_xS_y}-\frac{R_{xy}}{a^2}\right] \\ & +\frac{Q_y+R_{xy}S_y(b_y-Q_x)/S_x}{a^2S_y^2(1-R_{xy}^2)} \\ & +\frac{Q_x}{a^2S_x^2(1-R_{xy}^2)} \end{aligned}\right\}. \tag{7.8.5}$$

Hence, the posterior distribution of $\tilde{\mu}_y$ is normal with mean given in equation (7.8.5) and variance given in equations (7.8.3) and (7.8.4).

Assuming still the quadratic loss function as given in equation (7.5.20), the best estimator of μ_y is the posterior mean in equation (7.8.5). Note that $\tilde{\mu}_y = E(\tilde{\mu}_y \mid Q_x, Q_y)$ is a weighted average of the estimators b_y, Q_y, Q_x, and a regression of Q_y upon Q_x. As α_y^2 increases, the weight on b_y decreases. Likewise, as S_y^2 increases, the weight on Q_y decreases. The weight on Q_x decreases as both a^2 and S_x^2 increase. If S_x^2 is large, then Q_x is not a reliable estimator of μ_x; and if a^2 is large, then it is unlikely that $\mu_x = \mu_y$ and thus Q_x again is not a reliable estimator of μ_y. As S_y^2 increases, the weight on the regression component decreases because Q_y is an uncertain estimator of μ_y and it is not advantageous to correct Q_y by a regression upon Q_x. Also, the weight on the regression component decreases as R_{xy}, the correlation between $\tilde{Q}_x$ and $\tilde{Q}_y$, approaches zero since regression is not very meaningful if $\tilde{Q}_x$ and $\tilde{Q}_y$ are not correlated very much. Finally, the weight on the regression component decreases as a^2 increases since the adjustment of Q_y based on the relationship $b_y = b_x$ (i.e. b_y is assuming the position of μ_x) is hazardous.

The posterior loss in this section is

$$E\left[\text{Loss}\,(\tilde{\mu}_y, \hat{\mu}_y \mid Q_x, Q_y)\right] = bV(\tilde{\mu}_y \mid Q_x, Q_y) + nC_{xy} + n_x C_x + n_y C_y + C^*, \tag{7.8.6}$$

where $V(\tilde{\mu}_y \mid Q_x, Q_y)$ is given in equation (7.8.3). An optimum sample allocation to minimize the posterior loss for a fixed sampling budget C_o encounters the same difficulties as those mentioned in section E of this chapter. The possible solution suggested there would also be applicable here.

3. $\underline{\hat{\mu}_y}$ under limiting conditions of a^2

Consider now the estimator $\hat{\mu}_y$ when $a^2 \rightarrow \infty$. Letting $a^2 \rightarrow \infty$ is equivalent to giving a very small prior probability to the event that $\tilde{\Delta}$ is small. From equations (7.8.3) through (7.8.5), the limiting posterior mean and variance of $\hat{\mu}_y$ are

$$\lim_{a^2 \rightarrow \infty} E(\tilde{\mu}_y \mid Q_x, Q_y) = \frac{b_y/\alpha_y^2 + Q_y/s_y^2}{1/\alpha_y^2 + 1/s_y^2} \tag{7.8.7}$$

and

$$\lim_{a^2 \rightarrow \infty} V(\tilde{\mu}_y \mid Q_x, Q_y) = \left[1/\alpha_y^2 + 1/s_y^2\right]^{-1}. \tag{7.8.8}$$

Note that Q_x, which is an estimator of μ_x, appears nowhere in the estimator of equation (7.8.7). Recall, however, that Q_y, given in equation (7.5.1), does use the observations on the random variable $\tilde{x}$ in the form of a regression of $\bar{y}_n$ upon $\bar{x}_n$.

Letting $a^2 \rightarrow 0$ is equivalent to letting the prior probability that $\tilde{\mu}_y$ and $\tilde{\mu}_x$ are equal approach one. From equations (7.8.3) through (7.8.5), the limiting posterior mean and variance of $\tilde{\mu}_y$ can be obtained as

$$\lim_{a^2 \to 0} E(\tilde{\mu}_y \mid Q_x, Q_y) =$$

$$\left\{\lim_{a^2 \to 0} V(\tilde{\mu}_y \mid Q_x, Q_y)\right\}^{-1} \left[\frac{b_y}{\alpha_y^2} + \frac{Q_x}{(1-R_{xy}^2)} \left\{\frac{1}{S_x^2} - \frac{R_{xy}}{S_x S_y}\right\} + \frac{Q_y}{(1-R_{xy}^2)} \left\{\frac{1}{S_y^2} - \frac{R_{xy}}{S_x S_y}\right\} \right] \quad (7.8.9)$$

and

$$\lim_{a^2 \to 0} V(\tilde{\mu}_y \mid Q_x, Q_y) = \left\{\frac{1}{\alpha_y^2} + \frac{1}{S_y^2(1-R_{xy}^2)} + \frac{1}{S_x^2(1-R_{xy}^2)} - \frac{2R_{xy}}{S_x S_y (1-R_{xy}^2)}\right\}^{-1}. \quad (7.8.10)$$

Note that the estimator of μ_y in equation (7.8.9) is a weighted average of all information on μ_y and μ_x. It can be shown that the weights on b_y, Q_x, and Q_y in equation (7.8.9) minimize the "variance" of $(w_1 Q_x + w_2 Q_y + w_3 b_y)$, where $(w_1 + w_2 + w_3) = 1$ and expectation is taken with respect to the conditional distribution of $(\tilde{x}, \tilde{y})$, given (μ_x, μ_y). Also, b_y is considered to have "variance" α_y^2 and is assumed to be "independent" of $\tilde{Q}_x$ and $\tilde{Q}_y$. In equation (7.8.9) note that each of the three components is weighted, approximately, inversely proportional to its variance, where again α_y^2 is interpreted as the "variance" of b_y.

The posterior distribution of $\tilde{\mu}_y$ when $a^2 = 0$ can also be obtained by considering $\tilde{x}$ and $\tilde{y}$ to be bivariate normal with one unknown expectation $E(\tilde{x}) = E(\tilde{y}) = \mu_y$. The prior distribution on $\tilde{\mu}_y$ is then taken to be

$N(b_y, \alpha_y^2)$. This approach was discussed by Geisser (1965a) in his estimation of the mean of a multivariate normal distribution with unknown covariance matrix.

4. Special case where $\tilde{x}$ and $\tilde{y}$ are independent

Consider now the important special case where $\tilde{x}$ and $\tilde{y}$ are independently normally distributed and the prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ are also independent and normal. The posterior mean and variance of $\tilde{\mu}_y$ for this special case are obtained from equations (7.8.3) through (7.8.5) by letting $\rho = 0$ and $n = 0$. Recall that, when $\tilde{x}$ and $\tilde{y}$ are independent and $n = 0$,

$$Q_y = \bar{y}_{n_y}$$

$$Q_x = \bar{x}_{n_x}$$

$$s_y^2 = \sigma_y^2/n_y \tag{7.8.11}$$

$$s_x^2 = \sigma_x^2/n_x$$

$$R_{xy} = 0.$$

Thus, the posterior mean and variance are obtained as

$$E(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \frac{\left\{\dfrac{n_y\bar{y}_{n_y}}{\sigma_y^2} + \dfrac{b_y}{\alpha_y^2} + \dfrac{\bar{x}_{n_x}}{a^2+\sigma_x^2/n_x}\right\}}{\left\{\dfrac{n_y}{\sigma_y^2} + \dfrac{1}{\alpha_y^2} + \dfrac{1}{a^2+\sigma_x^2/n_x}\right\}} \tag{7.8.12}$$

and

$$V(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \left\{ \frac{n_y}{\sigma_y^2} + \frac{1}{\alpha_y^2} + \frac{1}{a^2 + \sigma_x^2/n_x} \right\}^{-1}. \tag{7.8.13}$$

Note that the posterior mean is a weighted average of three independent estimators of μ_y, i.e. $\bar{y}_{n_y}$, $\bar{x}_{n_x}$, and b_y. Note also, with the assumption of independence of $\tilde{x}$ and $\tilde{y}$, that the posterior mean no longer contains any regression estimators.

If a^2 is large, the weight on $\bar{x}_{n_x}$ is small, reflecting the uncertainty that $\Delta = 0$. The limiting posterior mean and variance of $\tilde{\mu}_y$ as $a^2 \longrightarrow \infty$ are

$$\lim_{a^2 \to \infty} E(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \frac{\left[b_y/\alpha_y^2 + n_y \bar{y}_{n_y}/\sigma_y^2 \right]}{\left[1/\alpha_y^2 + n_y/\sigma_y^2 \right]} \tag{7.8.14}$$

and

$$\lim_{a^2 \to \infty} V(\tilde{\mu}_y \mid \bar{x}_{n_y}, \bar{y}_{n_y}) = \left\{ 1/\alpha_y^2 + n_y/\sigma_y^2 \right\}^{-1}. \tag{7.8.15}$$

Hence, in this limiting case, none of the observations on $\tilde{x}$ are used in the estimator of μ_y. If, in addition, $\alpha_y^2 \longrightarrow \infty$, then the resulting estimator in equation (7.8.14) is similar to the preliminary test estimator in Chapter III when $H_A: \mu_y \neq \mu_x$ is accepted.

As $a^2 \longrightarrow 0$, the limiting posterior mean and variance of $\tilde{\mu}_y$ are

$$\lim_{a^2 \to 0} E(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \frac{\left[n_y \bar{y}_{n_y}/\sigma_y^2 + n_x \bar{x}_{n_x}/\sigma_x^2 + b_y/\alpha_y^2 \right]}{\left[n_y/\sigma_y^2 + n_x/\sigma_x^2 + 1/\alpha_y^2 \right]} \tag{7.8.16}$$

and

$$\lim_{a^2 \to 0} V(\tilde{\mu}_y \mid \bar{x}_{n_x}, \bar{y}_{n_y}) = \left\{ n_y/\sigma_y^2 + n_x/\sigma_x^2 + 1/\alpha_y^2 \right\}^{-1} . \qquad (7.8.17)$$

Again, if $\alpha_y^2 \to \infty$, the resulting estimator in equation (7.8.14) is similar to the preliminary test estimator in Chapter III when H_o: $\mu_y = \mu_x$ is accepted.

The optimum sample allocation of n_x and n_y which minimizes the posterior quadratic loss, given in equation (7.7.5), for a fixed sampling cost C_o is obtained from equations (7.7.13) and (7.7.19) by letting $\gamma = \alpha_y/\alpha_x$ and $\alpha_x^2 = a^2 + \alpha_y^2$. The critical value n_{yc} is then given by

$$n_{yc} = \frac{C_o}{C_y} + \frac{\sigma_x \sigma_y \sqrt{C_x}}{a^2 \sqrt{C_y}} \left[\frac{\sigma_x \sqrt{C_x}}{\sigma_y \sqrt{C_y}} - 1 \right] . \qquad (7.8.18)$$

The optimal sample size n_y is:

$$\begin{aligned} n_y &= n_{yc} \text{ if } 0 \leq n_{yc} \leq C_o/C_y \\ n_y &= C_o/C_y \text{ if } n_{yc} > C_o/C_y \\ n_y &= 0 \text{ if } n_{yc} < 0 . \end{aligned} \qquad (7.8.19)$$

Once n_y is determined, then

$$n_x = (C_o - C_y n_y)/C_x . \qquad (7.8.20)$$

From equation (7.8.18), $n_x > 0$ if, and only if,

$$\sigma_x \sqrt{C_x} < \sigma_y \sqrt{C_y} . \qquad (7.8.21)$$

Thus, even though $C_x < C_y$, if σ_x is very large with respect to σ_y, then the optimum sampling plan will include no observations on $\tilde{x}$. Given that inequality (7.8.21) is satisfied, then the sample size n_x increases as a^2 decreases.

If $a^2 = 0$, then the optimum sample allocation for a fixed sampling budget C_o is to sample only $\tilde{y}$ if $\sigma_x \sqrt{C_x} \geq \sigma_y \sqrt{C_y}$ and to sample only $\tilde{x}$ if $\sigma_x \sqrt{C_x} < \sigma_y \sqrt{C_y}$.

I. Normal Prior on $\tilde{\Delta}$ and Precise Measurement on $\tilde{\mu}_y$

1. The prior distribution on $\tilde{\Delta}$ and $\tilde{\mu}_y$

As in section H the prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ are assumed to be independent, and $\tilde{\Delta}$ is assumed to be distributed $N(0, a^2)$. Instead of assuming $\tilde{\mu}_y$ to be normally distributed as in section H, however, the principle of precise measurement is now applied to $\tilde{\mu}_y$.

2. Extension of precise measurement

The literature on precise measurement, mentioned in section B of this chapter, deals only with one parameter distributions. Although it is easy to extend the mechanics of the method to more than one parameter, an interpretation and graphical illustration as in Figure 7.1 is not obvious. An attempt to explain what the mechanics of the procedure imply is given below.

Let Q_y and Q_Δ be two statistics which are sufficient for the posterior analysis of $\tilde{\mu}_y$ and $\tilde{\Delta}$. (Q_y and Q_Δ are derived later in Lemma 7.2.) Then, the joint posterior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ is given by

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)\ g_1(\Delta)\ g_2(\mu_y), \tag{7.9.1}$$

where $g_1(\Delta)$ and $g_2(\mu_y)$ are the independent prior distributions on $\tilde{\Delta}$ and $\tilde{\mu}_y$, respectively, and $f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)$ is the conditional distribution of the two sufficient statistics given the parameters μ_y and Δ. Now, equation (7.9.1) can be written as

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)\ g_2(\mu_y), \tag{7.9.2}$$

where $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ is the joint distribution of $\tilde{Q}_y$, $\tilde{Q}_\Delta$, and $\tilde{\Delta}$, given the parameter μ_y. Replacing $g_2(\mu_y)$ by a constant gives an approximation to the posterior distribution as

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_2(Q_y, Q_\Delta, \Delta \mid \mu_y) \tag{7.9.3}$$

or

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)\ g_1(\Delta), \tag{7.9.4}$$

assuming, of course, that $g_1(\Delta) = g_1(\Delta \mid \mu_y)$. In order for equations (7.9.3) and (7.9.4) to be good approximations, $g_2(\mu_y)$ must be fairly constant for those values of μ_y and Δ for which $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ is large. Q_y and Q_Δ can be thought of as the sample results at hand, i.e. they can have specified numerical values. The following discussion assumes a given Q_y and Q_Δ, but obviously the properties described below must hold for any likely values of Q_y and Q_Δ which will result from the sampling scheme.

For a graphical interpretation, consider a three dimensional space with two of the axes labeled μ_y and Δ. Then, for fixed, numerical values of Q_y and Q_Δ, the function $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ can be graphed against μ_y and Δ. Also, the function $g_2(\mu_y)$ can be graphed against μ_y: this will be a cylinder since $g_2(\mu_y)$ is not a function of Δ. Now, for regions of the (μ_y, Δ)-plane where $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ is large or moderate, the cylinder $g_2(\mu_y)$ should be about the same height from the (μ_y, Δ)-plane. In addition, for regions of the (μ_y, Δ) plane where $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ is small, the distance of $g_2(\mu_y)$ from the (μ_y, Δ)-plane must not be many times larger than it is in the region where $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ is larger. A similar interpretation of precise measurement holds if $\tilde{\mu}_y$ and $\tilde{\Delta}$ are dependent, except that $g_1(\Delta)$ in the preceding discussion is replaced by $g_1(\Delta \mid \mu_y)$.

The foregoing discussion can be extended to consider the case where the principle of precise measurement is applied to the joint density $g(\mu_y, \Delta)$. In this case, the resultant posterior distribution of $\tilde{\mu}_y$ and $\tilde{\Delta}$ can be approximately represented as

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_1(Q_y, Q_\Delta \mid \mu_y, \Delta). \tag{7.9.5}$$

The comments of the preceding paragraphs then apply with $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ replaced now by $f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)$ and $g_2(\mu_y)$ replaced by $g(\mu_y, \Delta)$. Recall that precise measurement on $(\tilde{\mu}_y, \tilde{\Delta})$ was used to derive the estimator Q_y in section E of this chapter.

Precise measurement can probably be used without too much concern over the form of the prior distribution if sample sizes are large. This is because the joint distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$, given μ_y and Δ, will be fairly concentrated over a small region of the (Q_y, Q_Δ)-plane since $\tilde{Q}_y$ and $\tilde{Q}_\Delta$ will have relatively small variances. Thus, $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ or $f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)$, graphed as a function of μ_y and Δ for the fixed sample results Q_y and Q_Δ, will be concentrated over a fairly small region of the (μ_y, Δ)-plane. Hence, the assumption that $g_2(\mu_y)$ or $g(\mu_y, \Delta)$ be relatively constant over the region where $f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)$ or $f_2(Q_y, Q_\Delta\ \Delta \mid \mu_y)$ is large is not very restrictive since the behavior of the prior distribution is being specified only for a small region of the (μ_y, Δ)-plane. On the other hand, if the sample sizes are small, then the distributions $f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)$ and $f_2(Q_y, Q_\Delta, \Delta \mid \mu_y)$ are spread out over a larger portion of the (μ_y, Δ)-plane, and more stringent requirements on $g(\mu_y, \Delta)$ or $g_2(\mu_y)$ must be fulfilled in order for precise measurement to be justified. It should be emphasized again that although this discussion assumes fixed sample results Q_y and Q_Δ, the various properties which are discussed should be valid for any likely sample values of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$ in order to justify the principle of precise measurement.

3. Sufficient statistics

Lemma 7.2 below gives the sufficient statistics Q_y and Q_Δ mentioned previously for the posterior analysis of $\tilde{\mu}_y$ and $\tilde{\Delta}$.

Lemma 7.2. Let the sample data $y_1, y_2, \ldots, y_{n+n_y}, x_1, x_2, \ldots, x_{n+n_x}$ be selected as in Lemma 7.1 of section E from a bivariate normal population with unknown mean vector and know covariance matrix. Then a set of sufficient statistics for the Bayesian posterior analysis of $\tilde{\mu}_y$ and $\tilde{\Delta}$ is Q_y and Q_Δ, where

$$Q_\Delta = Q_y - Q_x \tag{7.9.6}$$

and Q_y and Q_x are given by equation (7.5.1).

The proof to this lemma is not given in detail, but two methods of proving it are indicated. First, a proof very similar to Lemma 7.1 can be constructed where the kernel of the likelihood function is found as a function of μ_y and Δ. This kernel can be shown to depend upon the two statistics Q_y and Q_Δ, and thus Q_y and Q_Δ are sufficient for the posterior analysis of $\tilde{\mu}_y$ and Δ. Second, since Q_y and Q_x are jointly sufficient for (μ_y, μ_x), and since the transformation from (Q_y, Q_x) and (μ_y, μ_x) to (Q_y, Q_Δ) and (μ_y, Δ) is one-to-one, then it is easily shown that Q_y and Q_Δ are jointly sufficient for (μ_y, Δ).

Since Q_y and Q_Δ form a set of sufficient statistics for the posterior analysis of $\tilde{\mu}_y$ and $\tilde{\Delta}$, then it is necessary to consider only the joint distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$, given μ_y and Δ, rather than the joint distribution of all the sample data. By Lemma 3.1 the joint distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$, given μ_y and Δ, is bivariate normal. Using equations (7.5.15) and (7.5.16), the mean vector and covariance matrix of $(\tilde{Q}_y, \tilde{Q}_\Delta)$ are easily found to be

$$E\begin{bmatrix} \tilde{Q}_y & \mu_y \\ \tilde{Q}_x & \Delta \end{bmatrix} = \begin{bmatrix} \mu_y \\ \Delta \end{bmatrix} \tag{7.9.7}$$

and

$$V\begin{bmatrix} \tilde{Q}_y & \mu_y \\ \tilde{Q}_x & \Delta \end{bmatrix} = \begin{bmatrix} S_y^2 & S_{y\Delta}^2 \\ S_{y\Delta}^2 & S_\Delta^2 \end{bmatrix}, \tag{7.9.8}$$

where S_y^2 is defined in equation (7.5.2) and

$$\begin{aligned} S_\Delta^2 &= S_y^2 + S_x^2 - 2S_{xy}^2 \\ S_{y\Delta}^2 &= S_y^2 - S_{xy}^2 = R_{y\Delta} S_y S_\Delta \end{aligned} \tag{7.9.9}$$

S_{xy}^2 is defined in equation (7.5.3), S_x^2 is defined by analogy with S_y^2, and $R_{y\Delta}$ is the correlation of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$.

Q_y was discussed in detail in section E, the main point being that Q_y is a weighted average of $\bar{y}_n$, $\bar{y}_{n_y}$, and the regression of $\bar{y}_n$ upon $\bar{x}_n$. Using equation (7.5.1), $Q_\Delta = Q_y - Q_x$ can be written as

$$Q_\Delta = \frac{\left\{ \begin{aligned} & n^2(\bar{y}_n - \bar{x}_n) + n_x n_y(1-\rho^2)(\bar{y}_{n_y} - \bar{x}_{n_x}) \\ & + n n_x\left[\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x - \bar{x}_{n_x}\right] \\ & + n n_y\left[\bar{y}_{n_y} - \left\{\bar{x}_n + \rho\sigma_x(\bar{y}_{n_y} - \bar{y}_n)/\sigma_y\right\}\right] \end{aligned} \right\}}{n^2 + n_x n_y(1-\rho^2) + n n_x + n n_y}. \tag{7.9.10}$$

Q_Δ is composed of a weighted average of four estimators of Δ, where the weights, which sum to one, minimize the conditional variance, given μ_y and Δ, of $\tilde{Q}_\Delta$. The estimator $(\bar{y}_n - \bar{x}_n)$ estimates Δ from the correlated bivariate sample, whereas $(\bar{y}_{n_y} - \bar{x}_{n_y})$ estimates Δ from the independent observations on $\tilde{x}$ and $\tilde{y}$. The third estimator estimates μ_y by means of a regression of $\bar{y}_n$ upon $\bar{x}_n$ and then subtracts $\bar{x}_{n_x}$ to provide an estimate of Δ. Likewise, the fourth estimator estimates $-\mu_y$ by means of a regression of $\bar{x}_n$ upon $\bar{y}_n$ and then adds $\bar{y}_{n_y}$ to provide an estimate of Δ.

4. Posterior distribution of $\tilde{\mu}_y$

Applying the principle of precise measurement to $\tilde{\mu}_y$, then the posterior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ is given by equation (7.9.4), i.e.

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto f_1(Q_y, Q_\Delta \mid \mu_y, \Delta)\, g_1(\Delta). \qquad (7.9.11)$$

Now, the distribution of $(\tilde{Q}_y, \tilde{Q}_\Delta)$, given (μ_y, Δ), is bivariate normal with parameters given in equations (7.9.7) and (7.9.8), and the distribution of $\tilde{\Delta}$ is $N(0, a^2)$. Thus

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \propto \exp\left\{-J/\left[2(1-R_{y\Delta}^2)\right]\right\} \qquad (7.9.12)$$

where

$$J = \left\{\frac{\mu_y - Q_y}{S_y}\right\}^2 + \left\{\frac{\Delta - Q_\Delta}{S_\Delta}\right\}^2 + \frac{\Delta^2(1-R_{y\Delta}^2)}{a^2} - 2R_{y\Delta}\left\{\frac{\mu_y - Q_y}{S_y}\right\}\left\{\frac{\Delta - Q_\Delta}{S_\Delta}\right\}, \qquad (7.9.13)$$

and all of the above terms are defined in equations (7.9.9), (7.9.10), and (7.5.1) through (7.5.2).

Define now the terms Z_y, Z_Δ, V_y^2, V_Δ^2, and $r_{y\Delta}$, where all other quantities below have been defined in Lemmas 7.1 and 7.2.

$$Z_y = Q_y - R_{y\Delta} S_y S_\Delta^{-1} \left[1 + a^2/S_\Delta^2\right]^{-1} Q_\Delta \tag{7.9.14}$$

$$Z_\Delta = Q_\Delta \left[1 + S_\Delta^2/a^2\right]^{-1} \tag{7.9.15}$$

$$r_{y\Delta}^{\ 2} = R_{y\Delta}^{\ 2} \left[1 + S_\Delta^2(1 - R_{y\Delta}^{\ 2})/a^2\right]^{-1} \tag{7.9.16}$$

$$V_y^2 = S_y^2 \left[1 + S_\Delta^2(1 - R_{y\Delta}^{\ 2})/a^2\right] \left[1 + S_\Delta^2/a^2\right]^{-1} \tag{7.9.17}$$

$$V_\Delta^2 = \left[1/S_\Delta^2 + 1/a^2\right]^{-1} \tag{7.9.18}$$

$$V_{y\Delta}^{\ 2} = r_{y\Delta} V_y V_\Delta \tag{7.9.19}$$

Now, after considerable algebraic manipulation, the kernel of the posterior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ in equation (7.9.12) can be written as

$$h(\mu_y, \Delta \mid Q_y, Q_\Delta) \quad \exp\left\{-\frac{1}{2(1-r_{y\Delta}^{\ 2})}\right\} \left[\left\{\frac{\mu_y - Z_y}{V_y}\right\}^2 + \left\{\frac{\Delta - Z_\Delta}{V_\Delta}\right\}^2 - 2r_{y\Delta}\left\{\frac{\mu_y - Z_y}{V_y}\right\}\left\{\frac{\Delta - Z_\Delta}{V_\Delta}\right\} \right] \tag{7.9.20}$$

By inspection of the kernel in equation (7.9.20), the posterior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ is seen to be bivariate normal with mean

$$E\left[\begin{array}{c|c} \tilde{\mu}_y & Z_y \\ \tilde{\Delta} & Z_\Delta \end{array}\right] = \begin{bmatrix} Z_y \\ Z_\Delta \end{bmatrix} \tag{7.9.21}$$

and covariance matrix

$$V\left[\begin{array}{c|c} \tilde{\mu}_y & Z_y \\ \tilde{\Delta} & Z_\Delta \end{array}\right] = \begin{bmatrix} V_y^2 & V_{y\Delta}^{\ 2} \\ V_{y\Delta}^{\ 2} & V_\Delta^2 \end{bmatrix}. \tag{7.9.22}$$

Thus, the posterior distribution of $\tilde{\mu}_y$ is $N(Z_y, V_y^2)$.

Recall now the posterior distribution of $\tilde{\mu}_y$ when the independent prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ are $N(b_y, \alpha_y^2)$ and $N(0, a^2)$, respectively. The posterior mean and variance of $\tilde{\mu}_y$ in this case, denoted by $E(\tilde{\mu}_y \mid Q_x, Q_y)$ and $V(\tilde{\mu}_y \mid Q_x, Q_y)$, are given by equations (7.8.5) and (7.8.3). Now, with a moderate amount of algebra and substitutions of the type $S_{y\Delta}^{\ 2} = S_y^2 - S_{xy}^{\ 2}$, etc. from equations (7.9.8) through (7.9.9), it can be shown that

$$\lim_{\alpha_y^2 \to \infty} E(\tilde{\mu}_y \mid Q_x, Q_y) = Z_y \tag{7.9.23}$$

and

$$\lim_{\alpha_y^2 \to \infty} V(\tilde{\mu}_y \mid Q_x, Q_y) = V_y^2. \tag{7.9.24}$$

Thus, taking a normal prior on $\tilde{\mu}_y$ with arbitrary constant mean and variance approaching infinity yields the same posterior distribution as applying the principle of precise measurement to $\tilde{\mu}_y$. $\tilde{\Delta}$, in both cases, is $N(0, a^2)$ and independent of $\tilde{\mu}_y$. Note also that a normal distribution with variance approaching infinity could be interpreted as a uniform distribution on the real line, a concept discussed by Jeffreys (1961).

5. Estimator of μ_y

Using the quadratic loss function in equation (7.5.20), the best estimator of μ_y is the posterior mean of $\tilde{\mu}_y$ given in equation (7.9.14), i.e.

$$E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = Q_y - R_{y\Delta} S_y S_\Delta^{-1} (1 + a^2/S_\Delta^2)^{-1} \, Q_\Delta = Z_y. \qquad (7.9.25)$$

Note that Z_y is a weighted average of Q_y and Q_Δ, i.e. a weighted average of Q_y and Q_x since $Q_\Delta = Q_y - Q_x$. It is easy to show that Z_y can also be written as

$$Z_y = \frac{\dfrac{Q_y}{S_\Delta^2} + \dfrac{1}{a^2}\left\{Q_y + \dfrac{R_{y\Delta} S_y (0 - Q_\Delta)}{S_\Delta}\right\}}{\left\{\dfrac{1}{S_\Delta^2} + \dfrac{1}{a^2}\right\}}. \qquad (7.9.26)$$

In this form Z_y is thus a weighted average of the estimator Q_y and the regression of Q_y upon Q_Δ, where the prior mean of $\tilde{\Delta}$, i.e. $E(\tilde{\Delta}) = 0$, is in the position usually occupied by the expected value of $\tilde{Q}_\Delta$. Recall that $E(\tilde{Q}_\Delta \mid \mu_y, \Delta) = \Delta$. Regarding Z_y as in equation (7.9.26), the weight on the regression component decreases as a^2 increases. This occurs because a large a^2 implies that the value zero is an uncertain estimate of $E(\tilde{Q}_\Delta \mid \mu_y, \Delta)$, and hence the regression of Q_y upon Q_Δ is hazardous. Also, as S_Δ^2 increases, Q_y is weighted more heavily and the regression term becomes negligible. This occurs because, with a large S_Δ^2, Q_Δ is an uncertain estimate of Δ, and again regression is risky. Recall that Q_y itself is a weighted average of $\bar{y}_n$, $\bar{y}_{n_y}$, and the regression of $\bar{y}_n$ upon $\bar{x}_n$.

The posterior loss for the estimator $\hat{\mu}_y = Z_y$ is

$$E\left[\text{Loss}(\tilde{\mu}_y, \hat{\mu}_y \mid Q_y, Q_\Delta)\right] = bV(\tilde{\mu}_y \mid Q_y, Q_\Delta) + nC_{xy} + n_x C_x + n_y C_y + C^* \tag{7.9.27}$$

where

$$V(\tilde{\mu}_y \mid Q_y, Q_\Delta) = S_y^2 \left[1 + S_\Delta^2(1-R_{y\Delta}^2)/a^2\right] \left[1+S_\Delta^2/a^2\right]^{-1}$$

$$= V_y^2. \tag{7.9.28}$$

An attempt to find a sample allocation to minimize the posterior loss for a fixed sampling budget C_o encounters the same difficulties mentioned earlier in section E of this chapter.

It is of interest to examine the estimator Z_y under limiting conditions of a^2. If $a^2 \rightarrow \infty$, then

$$\lim_{a^2 \rightarrow \infty} Z_y = Q_y \tag{7.9.29}$$

and

$$\lim_{a^2 \rightarrow \infty} V_y^2 = S_y^2. \tag{7.9.30}$$

Thus, the estimator of μ_y is simply Q_y, as in section E where the prior distribution of $(\tilde{\mu}_y, \tilde{\mu}_x)$ was taken to be a constant. These two procedures

- i.e. 1) using precise measurement on $\tilde{\mu}_y$ and assigning an independent $N(0, a^2)$ distribution to $\tilde{\Delta}$ with infinite variance a^2, and 2) regarding the prior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ to be a constant - can be considered as two attempts to incorporate the idea of no prior information into the Bayesian framework. Both attempts give Q_y as the posterior mean, and thus both procedures use the auxiliary information on $\tilde{x}$ in the form of a regression of $\bar{y}_n$ upon $\bar{x}_n$.

If $a^2 \longrightarrow 0$, then

$$\lim_{a^2 \to 0} Z_y = Q_y - R_{y\Delta} S_y Q_\Delta / S_\Delta$$

$$= \frac{Q_y(S_x^2 - S_{xy}^2) + Q_x(S_y^2 - S_{xy}^2)}{(S_x^2 + S_y^2 - 2S_{xy}^2)} \tag{7.9.31}$$

and

$$\lim_{a^2 \to 0} V_y^2 = S_y^2(1 - R_{y\Delta}^2). \tag{7.9.32}$$

In this case, $\hat{\mu}_y$ in equation (7.9.31) is a weighted average of Q_x and Q_y. It can be shown that the weights in equation (7.9.31) minimize $V\left[w_1\tilde{Q}_y + w_2\tilde{Q}_x\right]$ where $\left[w_1 + w_2\right] = 1$ and expectation is taken with respect to the conditional distribution of $(\tilde{Q}_y, \tilde{Q}_x)$, given (μ_y, Δ) or (μ_y, μ_x). Note that the estimator in equation (7.9.31) allows for no uncertainty that Q_x actually is estimating μ_y, since it was obtained by letting $a^2 \longrightarrow 0$.

6. Special case where $\tilde{x}$ and $\tilde{y}$ are independent

Let $\tilde{x}$ and $\tilde{y}$ be independently normally distributed, and let the prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ be independent, where $\tilde{\Delta}$ is $N(0, a^2)$ and precise measurement is applied to $\tilde{\mu}_y$. Then the posterior mean and variance of $\tilde{\mu}_y$ are given by equations (7.9.14) and (7.9.17) with ρ and n equated to zero. Thus, for $\tilde{x}$ and $\tilde{y}$ independent,

$$E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = V\left[\tilde{\mu}_y \mid Q_y, Q_\Delta\right] \left[\frac{n_y \bar{y}_{n_y}}{\sigma_y^2} + \frac{\bar{x}_{n_x}}{a^2 + \sigma_x^2/n_x}\right] \tag{7.9.33}$$

and

$$V(\tilde{\mu}_y \mid Q_y, Q_\Delta) = \left[\frac{n_y}{\sigma_y^2} + \frac{1}{a^2 + \sigma_x^2/n_x}\right]^{-1}. \tag{7.9.34}$$

Thus, $\hat{\mu}_y = E(\tilde{\mu}_y \mid Q_y, Q_\Delta)$ is a weighted average of the two independent estimates $\bar{y}_{n_y}$ and $\bar{x}_{n_x}$. Note that the weight on $\bar{y}_{n_y}$ is inversely proportional to $V(\tilde{\bar{y}}_{n_y} \mid \mu_y, \Delta)$. Also, the weight on $\bar{x}_{n_x}$ is inversely proportional to the sum of the two variances $V(\tilde{\Delta})$ and $V(\tilde{\bar{x}}_{n_x} \mid \mu_y, \Delta)$. As a^2 increases, the component $\bar{x}_{n_x}$ is weighted less.

If $a^2 \to 0$, then

$$\lim_{a^2 \to 0} E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = \left\{\frac{n_y \bar{y}_{n_y}}{\sigma_y^2} + \frac{n_x \bar{x}_{n_x}}{\sigma_x^2}\right\} \left\{\frac{n_y}{\sigma_y^2} + \frac{n_x}{\sigma_x^2}\right\}^{-1} \tag{7.9.35}$$

which is the pooled estimator in the classical preliminary test approach when $H_o: \mu_y = \mu_x$ is accepted. Also, if $a^2 \to \infty$, then

$$\lim_{a^2 \to \infty} E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = \bar{y}_{n_y}, \tag{7.9.36}$$

which is the estimator in the classical preliminary test approach when H_A: $\mu_y \neq \mu_x$ is accepted.

The optimum sample allocation for n_x and n_y with the quadratic loss function in equation (7.7.5) and fixed sampling budget C_o is given by equations (7.8.18) through (7.8.20).

J. Normal Prior on $\tilde{\Delta}$ and Uniform Distribution on $\tilde{\mu}_y$

1. Posterior distribution of $\tilde{\mu}_y$

Let $\tilde{x}$, $\tilde{y}$, and $\tilde{\Delta}$ have the same distributions as in section F. Now, however, let $\tilde{\mu}_y$ have a uniform distribution over the interval $[c_1, c_2]$, where $\tilde{\mu}_y$ is still assumed to be independent of $\tilde{\Delta}$, and $\tilde{\Delta}$ is $N(0, a^2)$. The prior density of $\tilde{\mu}_y$ is a constant, and thus the joint posterior distribution of $\tilde{\mu}_y$ and $\tilde{\Delta}$ is proportional to the product of the kernels of the likelihood and the prior distribution on $\tilde{\Delta}$. This same argument was used in section I, so equation (7.9.20) gives the kernel of the posterior distribution of $\tilde{\mu}_y$ and $\tilde{\Delta}$ for this case also. However, the posterior distribution is not defined over the (μ_y, Δ)-plane as in the previous section, but over the region $c_1 \leq \mu_y \leq c_2$ and $-\infty < \Delta < \infty$. Thus, the joint posterior distribution is not bivariate normal.

Integrating the kernel in equation (7.9.20) with respect to Δ over the real line yields the posterior distribution of $\tilde{\mu}_y$ as

$$h(\mu_y \mid Q_y, Q_\Delta) \propto \exp - \left\{ \frac{(\mu_y - Z_y)^2}{2V_y^2} \right\}, \tag{7.10.1}$$

where V_y^2 and Z_y are defined in equations (7.9.17) and (7.9.14). Since $h(\mu_y \mid Q_y, Q_\Delta)$ is non-zero only for $c_1 \leq \mu_y \leq c_2$, then $h(\mu_y \mid Q_y, Q_\Delta)$ is a truncated normal distribution with density

$$h(\mu_y \mid Q_y, Q_\Delta) = \frac{\exp\left\{ - \dfrac{(\mu_y - Z_y)^2}{2V_y^2} \right\}}{\sqrt{2\pi}\, V_y \left[\Phi\left\{ \dfrac{c_2 - Z_y}{V_y} \right\} - \Phi\left\{ \dfrac{c_1 - Z_y}{V_y} \right\} \right]}, \tag{7.10.2}$$

where $\Phi(x)$ is the cumulative normal distribution function defined in equation (3.7.34). As $c_1 \longrightarrow -\infty$ and $c_2 \longrightarrow \infty$, i.e. a limiting uniform distribution on the real line,

$$\lim_{\substack{c_1 \to \infty \\ c_2 \to \infty}} h(\mu_y \mid Q_y, Q_\Delta) = \frac{\exp\left\{ - \dfrac{(\mu_y - Z_y)^2}{2V_y^2} \right\}}{\sqrt{2\pi}\, V_y}. \tag{7.10.3}$$

Thus, as the length of the interval $[c_1, c_2]$ increases, the posterior truncated normal distribution approaches the posterior distribution obtained in section I by applying the principle of precise measurement to $\tilde{\mu}_y$.

Letting

$$G_1 = (c_1 - Z_y)/V_y$$
$$G_2 = (c_2 - Z_y)/V_y, \tag{7.10.4}$$

the posterior mean of $\tilde{\mu}_y$ is obtained as

$$E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = Z_y - \frac{V_y\left[\phi(G_2) - \phi(G_1)\right]}{\left[\Phi(G_2) - \Phi(G_1)\right]}, \tag{7.10.5}$$

where $\phi(x)$ is the N(0, 1) density defined in equation (3.7.4). Thus, the estimator $\hat{\mu}_y = E(\tilde{\mu}_y \mid Q_y, Q_\Delta)$ is composed of Z_y plus a correction factor C, where

$$C = \frac{-V_y\left[\phi(G_2) - \phi(G_1)\right]}{\left[\Phi(G_2) - \Phi(G_1)\right]}. \tag{7.10.6}$$

The correction factor brings $\hat{\mu}_y$ into the range $c_1 \leq \hat{\mu}_y \leq c_2$ even though Z_y may not be in this range. The correction term is zero if, and only if,

$$G_1 = -G_2, \tag{7.10.7}$$

i.e. if, and only if,

$$Z_y = (c_1 + c_2)/2 \tag{7.10.8}$$

where $(c_1 + c_2)/2$ is the mean of the prior distribution on $\tilde{\mu}_y$. If $Z_y \neq (c_1 + c_2)/2$, then it is easily shown that the estimator $\tilde{\mu}_y$ corrects Z_y toward the prior mean $(c_1 + c_2)/2$.

An approximation to the correction factor in equation (7.10.6) can be obtained by expanding $\phi(G_2)$, $\phi(G_1)$, $\Phi(G_2)$, and $\Phi(G_1)$ in a Taylor series around the point m, where m, the midpoint of the interval (G_1, G_2), is

$$m = (G_1 + G_2)/2 = V_y^{-1}\left[(c_1 + c_2)/2 - Z_y\right] . \tag{7.10.9}$$

Retaining only the terms of the series of degree four or less in m, the approximation is

$$\frac{-V_y\left[\phi(G_2) - \phi(G_1)\right]}{\Phi(G_2) - \Phi(G_1)} \doteq \frac{mV_y\left[1 + \frac{k^2(m^2-3)}{3!}\right]}{\left[1 + \frac{k^2(m^2-1)}{3!}\right]} \tag{7.10.10}$$

where

$$k = (c_2 - c_1)/2V_y. \tag{7.10.11}$$

Note that if m = 0, i.e. $Z_y = (c_1 + c_2)/2$, then both the correction term and the approximation are zero, resulting in the posterior mean being $Z_y = (c_1 + c_2)/2$. Also, as $k \longrightarrow 0$, then the approximation to the correction term approaches $(c_1 + c_2)/2 - Z_y$; thus the posterior mean approaches $(c_1 + c_2)/2$. This reflects the fact that when the prior is very informative, the posterior mean is weighted much more heavily by the prior information than by the sample data.

$E(\tilde{\mu}_y^2 \mid Q_y, Q_\Delta)$ is evaluated by using integration by parts, and the posterior variance of $\tilde{\mu}_y$, after some calculus, is obtained as

$$V(\tilde{\mu}_y \mid Q_y, Q_\Delta) = V_y^2\left\{1 - \frac{\left[G_2\phi(G_2) - G_1\phi(G_1)\right]}{\left[\Phi(G_2) - \Phi(G_1)\right]} - \frac{\left[\phi(G_2) - \phi(G_1)\right]^2}{\left[\Phi(G_2) - \Phi(G_1)\right]^2}\right\} . \tag{7.10.12}$$

2. <u>The loss function and unconditional distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$</u>

Using the quadratic loss function specified in equation (7.5.20), the posterior loss is

$$E\left[L(\tilde{\mu}_y, \hat{\mu}_y \mid Q_y, Q_\Delta)\right] = bV(\tilde{\mu}_y \mid Q_y, Q_\Delta) + nC_{xy} + n_x C_x + n_y C_y + C^* \tag{7.10.13}$$

where $V(\tilde{\mu}_y \mid Q_y, Q_\Delta)$ is given in equation (7.10.12). Note that the posterior loss is a function of the sample values Q_y and Q_Δ, since G_1 and G_2 are both functions of Z_y. In order to investigate the magnitude of the loss function before the sample data are collected, the expectation of equation (7.10.13) is taken over the unconditional joint distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$. This yields the unconditional expected loss.

The conditional distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$, given μ_y and Δ, is bivariate normal with mean and covariance matrix given in equations (7.9.7) and (7.9.8). Then the joint distribution of $\tilde{Q}_y$, $\tilde{Q}_\Delta$, $\tilde{\mu}_y$, and $\tilde{\Delta}$ is $h(Q_y, Q_\Delta, \mu_y, \Delta)$, where

$$h(Q_y, Q_\Delta, \mu_y, \Delta) = K^{-1} \exp\left\{-\frac{J}{2(1-R_{y\Delta}^2)}\right\}. \tag{7.10.14}$$

J is given in equation (7.9.13), and K is easily obtained as

$$K = (2\pi)^{3/2}\, a(c_2-c_1)\, S_y S_\Delta \sqrt{1-R_{y\Delta}^2}\ . \tag{7.10.15}$$

Using equations (7.9.14) through (7.9.19) it can be shown that

$$h(Q_y, Q_\Delta, \mu_y, \Delta) = K^{-1} \exp \left\{ \frac{-Q_\Delta^2 V_\Delta^2}{2a^2 S_\Delta^2} \right\}$$

$$\times \exp \left\{ -\frac{1}{2(1-r_{y\Delta}^2)} \right\} \left[\left\{ \frac{\mu_y - Z_y}{V_y} \right\}^2 + \left\{ \frac{\Delta - Z_\Delta}{V_\Delta} \right\}^2 - 2r_{y\Delta} \left\{ \frac{\mu_y - Z_y}{V_y} \right\} \left\{ \frac{\Delta - Z_\Delta}{V_\Delta} \right\} \right]. \tag{7.10.16}$$

Integrating equation (7.10.16) with respect to Δ over the real line yields the joint distribution of $\tilde{Q}_y$, $\tilde{Q}_\Delta$, and $\tilde{\mu}_y$ as

$$h(Q_y, Q_\Delta, \mu_y) = K^{-1}\sqrt{2\pi}\, V_\Delta \sqrt{1-r_{y\Delta}^2}$$

$$\times \exp \left\{ \frac{-Q_\Delta^2 V_\Delta^2}{2a^2 S_\Delta^2} - \frac{(\mu_y - Z_y)^2}{2V_y^2} \right\}. \tag{7.10.17}$$

Integrating equation (7.10.17) with respect to μ_y over the interval $[c_1, c_2]$ yields the unconditional distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$ as

$$h(Q_y, Q_\Delta) = \frac{\left[\Phi(G_2) - \Phi(G_1)\right] V_\Delta}{\sqrt{2\pi}\,(c_2 - c_1) a S_\Delta} \exp \left\{ \frac{-Q_\Delta^2 V_\Delta^2}{2a^2 S_\Delta^2} \right\}, \tag{7.10.18}$$

where G_1 and G_2 are defined in equation (7.10.4).

To confirm that the integral of equation (7.10.18) over the (Q_y, Q_Δ)-plane equals one, make the change of variable

$$\begin{aligned} y &= Z_y = Q_y - Q_\Delta\, r_{y\Delta}\, V_y\, V_\Delta / a^2 \\ z &= -Q_\Delta\, r_{y\Delta}\, V_y\, V_\Delta / a^2 \end{aligned} \tag{7.10.19}$$

and integrate first with respect to z, which yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(Q_y, Q_\Delta)\, dQ_y\, dQ_\Delta = \frac{1}{(c_2-c_1)} \int_{-\infty}^{\infty} \left[\Phi\left\{\frac{c_2-y}{V_y}\right\} - \Phi\left\{\frac{c_1-y}{V_y}\right\}\right] dy. \tag{7.10.20}$$

The final integration is evaluated by substituting

$$\Phi\left\{\frac{c_2-y}{V_y}\right\} - \Phi\left\{\frac{c_1-y}{V_y}\right\} = (2\pi)^{-\frac{1}{2}} \int_{\frac{c_1-y}{V_y}}^{\frac{c_2-y}{V_y}} e^{-t^2/2}\, dt \tag{7.10.21}$$

and then reversing the order of integration.

In order to find the unconditional posterior loss, it is necessary to evaluate $E\left[V(\tilde{\mu}_y \mid Q_y, Q_\Delta)\right]$ in equation (7.10.12), where expectation is taken over the unconditional distribution of $(\tilde{Q}_y, \tilde{Q}\ )$ given in equation (7.10.18). Proceeding now to evaluate $E\left[V(\tilde{\mu}_y \mid Q_y, Q_\Delta)\right]$, the expectation of the first term is

$$E\left[\frac{G_2\phi(G_2) - G_1\phi(G_1)}{\Phi(G_2) - \Phi(G_1)}\right] = E_1$$

$$= \frac{V_\Delta}{\sqrt{2\pi}\,(c_2-c_1)aS_\Delta} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[G_2\phi(G_2) - G_1\phi(G_1)\right] \exp\left\{\frac{-Q_\Delta^2 V_\Delta^2}{2a^2 S_\Delta^2}\right\} dQ_y\, dQ_\Delta. \tag{7.10.22}$$

Making the same transformation as in equation (7.10.19) and integrating with respect to z yields

$$E_1 = \frac{1}{(c_2-c_1)} \int_{-\infty}^{\infty} \left[\left\{\frac{c_2-y}{V_y}\right\} \phi \left\{\frac{c_2-y}{V_y}\right\} - \left\{\frac{c_1-y}{V_y}\right\} \phi \left\{\frac{c_1-y}{V_y}\right\} \right] dy, \tag{7.10.23}$$

which implies

$$E\left[\frac{G_2\phi(G_2) - G_1\phi(G_1)}{\Phi(G_2) - \Phi(G_1)}\right] = 0. \tag{7.10.24}$$

The second expectation, i.e.

$$E\left[\frac{\phi(G_2) - \phi(G_1)}{\Phi(G_2) - \Phi(G_1)}\right]^2 = E_2, \tag{7.10.25}$$

cannot be evaluated as neatly, but obviously it is positive. Thus,

$$E\left[V(\tilde{\mu}_y \mid Q_y, Q_\Delta)\right] < V_y^2. \tag{7.10.26}$$

Recall from the previous section, where precise measurement was applied to $\tilde{\mu}_y$, that the posterior variance $V(\tilde{\mu}_y \mid Q_y, Q_\Delta)$ equals V_y^2. Thus, its expectation over the unconditional distribution of $\tilde{Q}_y$ and $\tilde{Q}_\Delta$ also equals V_y^2. Hence, the expectation of the posterior variance for the case where $\tilde{\mu}_y$ has a uniform distribution over (c_1, c_2) is always less than the case where precise measurement is applied to $\tilde{\mu}_y$.

3. Special case where $\tilde{x}$ and $\tilde{y}$ are independent

If $\tilde{x}$ and $\tilde{y}$ are independent, then the posterior mean is

$$E(\tilde{\mu}_y \mid Q_y, Q_\Delta) = Z_y - \frac{V_y\left[\phi(G_2) - \phi(G_1)\right]}{\Phi(G_2) - \Phi(G_1)}, \tag{7.10.27}$$

where now

$$Z_y = V_y^2\left[\frac{n_y\bar{y}_{n_y}}{\sigma_y^2} + \frac{\bar{x}_{n_x}}{a^2+\sigma_x^2/n_x}\right], \tag{7.10.28}$$

$$V_y^2 = \left\{\frac{n_y}{\sigma_y^2} + \frac{1}{a^2+\sigma_x^2/n_x}\right\}^{-1}, \tag{7.10.29}$$

and G_1 and G_2 are defined in equation (7.10.4).

K. A Summary of the Estimators of μ_y

1. Sufficient statistics

In this chapter five different prior distributions are considered on either $(\tilde{\mu}_x, \tilde{\mu}_y)$ or on $(\tilde{\mu}_y, \tilde{\Delta})$: (1) precise measurement on the joint distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$, (2) bivariate normal prior on $(\tilde{\mu}_x, \tilde{\mu}_y)$, (3) independent normal priors on $\tilde{\mu}_y$ and $\tilde{\Delta}$, (4) normal prior on $\tilde{\Delta}$ and precise measurement on $\tilde{\mu}_y$, (5) normal prior on $\tilde{\Delta}$ and an independent uniform prior over (c_1, c_2) on $\tilde{\mu}_y$. For the estimation of μ_y, precise measurement on $(\tilde{\mu}_y, \tilde{\Delta})$ will yield the same estimator as procedure (1) above. Also, procedure (3) is derived as an important special case of procedure (2).

If the posterior analysis is done on $(\tilde{\mu}_x, \tilde{\mu}_y)$, then the statistics Q_x and Q_y are sufficient; likewise, Q_y and Q_Δ are sufficient for the posterior analysis on $(\tilde{\mu}_y, \tilde{\Delta})$, where

$$Q_y = \frac{nn_x\left[\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x\right] + n^2\bar{y}_n + n_y\left[n + n_x(1-\rho^2)\right]\bar{y}_{n_y}}{nn_x + n^2 + n_y\left[n_x(1-\rho^2) + n\right]}, \tag{7.11.1}$$

Q_x is obtained by substituting x for y in equation (7.11.1), and

$$Q_\Delta = Q_y - Q_x. \tag{7.11.2}$$

In general the main interest is on μ_y and thus the posterior distribution of $\tilde{\mu}_y$ is obtained. Note that Q_y is a weighted average of three estimators of μ_y: the mean $\bar{y}_n$, the mean $\bar{y}_{n_y}$, and the usual linear regression of $\bar{y}_n$ upon $\bar{x}_n$ with μ_x estimated by $\bar{x}_{n_x}$. A similar interpretation holds for Q_x and Q_Δ.

2. $\underline{\tilde{x}}$ and $\underline{\tilde{y}}$ dependent

If the principle of precise measurement is applied to the joint distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ or $(\tilde{\mu}_y, \tilde{\Delta})$, then the estimator of μ_y is Q_y, where Q_y is given in equation (7.11.1). If $n > 0$, $n_x > 0$, and $n_y > 0$, then Q_y utilizes all of the sample information. Note that both $\bar{x}_n$ and $\bar{x}_{n_x}$ are needed for the regression term. Thus, in order to use all of the available sample data, n_x and n must both be positive or both zero. (In the estimation scheme in this chapter, it is useless, e.g. to have $n_x = 0$ and $n > 0$

because the n observations $x_1, x_2, \ldots, x_n$ are not utilized in the estimator of μ_y.)

When a bivariate normal prior is taken on $(\tilde{\mu}_x, \tilde{\mu}_y)$, the estimator of μ_y, given by equations (7.6.13), (7.6.12), and (7.5.1) through (7.5.5), is a weighted average of b_y, Q_y, the regression of Q_y on Q_x where μ_x is replaced by b_x, and the regression of b_y upon b_x where μ_x is replaced by Q_x. For a discussion of the weights on the various components, see section F of this chapter. Since both Q_y and Q_x appear in the estimator, then all the observations on $\tilde{x}$ are utilized even if, for example, $n = 0$.

If the prior distributions on $\tilde{\mu}_y$ and $\tilde{\Delta}$ are independently normally distributed, then the estimator of μ_y, given by equations (7.8.5), (7.8.4), and (7.5.1) through (7.5.5), is a weighted average of b_y, Q_y, Q_x, and the regression of Q_y upon Q_x where μ_x is replaced by b_y. Note in the previous paragraph (where the prior on $(\tilde{\mu}_x \ \tilde{\mu}_y)$ is bivariate normal) that Q_x appears only in the two regression terms. However, in the case where $\tilde{\mu}_y$ and $\tilde{\Delta}$ are normally independently distributed, Q_x appears in the weighted average with a weight inversely proportional to a^2, where a^2 is the prior variance of $\tilde{\Delta}$. Also, in this case there is only one regression component. The regression of b_y upon b_x, which appeared in the estimator of the previous paragraph, does not appear here since $b_y = b_x$. Again, all data on $\tilde{x}$ is used in the estimator even if n and/or n_x are zero.

If now the prior on $\tilde{\Delta}$ is $N(0, a^2)$ and the principle of precise measurement is applied to $\tilde{\mu}_y$, then the estimator of μ_y is Z_y, where Z_y, given in equation (7.9.26), is

$$Z_y = \frac{\frac{Q_y}{S_\Delta^2} + \frac{1}{a^2}\left\{\left[Q_y + \frac{R_{y\Delta}S_y(0-Q_\Delta)}{S_\Delta}\right]\right\}}{\left\{\left[\frac{1}{S_\Delta^2} + \frac{1}{a^2}\right]\right\}} . \tag{7.11.3}$$

Note that Z_y is a weighted average of Q_y and the regression of Q_y upon Q_Δ. If a^2 is large, then Q_y receives the largest weight. Likewise, if S_Δ^2 is large, Q_y predominates. All sample information on $\tilde{x}$ is used in the estimator for any $n \geq 0$, $n_x \geq 0$, and $n_y \geq 0$.

If $\tilde{\Delta}$ is $N(0, a^2)$ and $\tilde{\mu}_y$ is independently distributed uniformly over (c_1, c_2), then the estimator of μ_y is

$$Z_y + C, \tag{7.11.4}$$

where Z_y is given in equation (7.11.3) and C is a correction factor given in equation (7.10.6) The correction factor C adjusts Z_y toward the prior mean $(c_1 + c_2)/2$. If $c_1 \longrightarrow -\infty$ and $c_2 \longrightarrow \infty$, then this case is the same as the previous one where precise measurement was applied to $\tilde{\mu}_y$. If, in addition, $a^2 \longrightarrow \infty$, then this case further reduces to the first case considered where precise measurement was applied to $(\tilde{\mu}_y, \tilde{\mu}_x)$ or $(\tilde{\mu}_y, \tilde{\Delta})$. In this case, also, all sample data on $\tilde{x}$ is used in the estimator for any $n \geq 0$, $n_x \geq 0$, $n_y \geq 0$.

3. $\tilde{x}$ and $\tilde{y}$ independent

If $\tilde{x}$ and $\tilde{y}$ are independent, then n is set equal to zero, and the sufficient statistics Q_x, Q_y, and Q_Δ become

$$\begin{aligned} Q_x &= \bar{x}_{n_x} \\ Q_y &= \bar{y}_{n_y} \\ Q_\Delta &= \bar{y}_{n_y} - \bar{x}_{n_x} . \end{aligned} \tag{7.11.5}$$

Of course, if the principle of precise measurement is applied to the joint distribution of $(\tilde{\mu}_x, \tilde{\mu}_y)$ or $(\tilde{\mu}_y, \tilde{\Delta})$, then the estimator of μ_y is simply $\bar{y}_{n_y}$.

If $(\tilde{\mu}_x, \tilde{\mu}_y)$ has a bivariate normal joint prior distribution, then the estimator of μ_y, given by equation (7.7.2), is

$$\frac{1}{D}\left[\frac{b_y}{\alpha_x^2\alpha_y^2(1-\gamma^2)} + \frac{\bar{y}_{n_y} n_y}{\sigma_y^2}\left\{\frac{n_x}{\sigma_x^2} + \frac{1}{\alpha_x^2(1-\gamma^2)}\right\} + \frac{n_x\left[b_y + \gamma\alpha_y(\bar{x}_{n_x} - b_x)/\alpha_x\right]}{\sigma_x^2\alpha_y^2(1-\gamma^2)}\right] \tag{7.11.6}$$

where the constant D is given in equation (7.7.4). The above estimator is a weighted average of b_y, $\bar{y}_{n_y}$, and the regression of b_y upon b_x. For a thorough discussion of the weights, see section G of this chapter.

If $\tilde{\mu}_y$ and $\tilde{\Delta}$ are independent normally distributed, then the estimator of μ_y is given by

$$\frac{\left[\frac{n_y \bar{y}_{n_y}}{\sigma_y^2} + \frac{b_y}{\sigma_y^2} + \frac{\bar{x}_{n_x}}{(a^2+\sigma_x^2/n)}\right]}{n_y/\sigma_y^2 + 1/\alpha_y^2 + 1/(a^2+\sigma_y^2/n)} \circ \tag{7.11.7}$$

This estimator is a weighted average of $\bar{y}_{n_y}$, b_y, and $\bar{x}_{n_x}$, where the weight on $\bar{x}_{n_x}$ is of $O(1/a^2)$.

If $\tilde{\Delta}$ is distributed $N(0, a^2)$ independently of $\tilde{\mu}_y$ and the principle of precise measurement is applied to $\tilde{\mu}_y$, then the estimator of μ_y is

$$\frac{\left[\frac{n_y \bar{y}_{n_y}}{\sigma_y^2} + \frac{\bar{x}_{n_x}}{(a^2+\sigma_x^2/n_x)}\right]}{n_y/\sigma_y^2 + 1/(a^2+\sigma_x^2/n_x)} \cdot \tag{7.11.8}$$

This can also be obtained from equation (7.11.7) by letting $\alpha_y^2 \longrightarrow \infty$.

If $\tilde{\Delta}$ is normally distributed and $\tilde{\mu}_y$ is independently uniformly distributed over (c_1, c_2), then the estimator of μ_y is

$$\frac{\left[\frac{n_y \bar{y}_{n_y}}{\sigma_y^2} + \frac{\bar{x}_{n_x}}{(a^2+\sigma_x^2/n_x)}\right]}{n_y/\sigma_y^2 + 1/(a^2+\sigma_x^2/n_x)} + C, \tag{7.11.9}$$

where the term C corrects the estimator toward the prior mean $(c_1 + c_2)/2$. C is given in equation (7.10.6).

VIII. A COMPARISON OF SOME BAYESIAN AND CLASSICAL ESTIMATORS

A. Sampling Scheme

In this chapter two Bayesian estimators and three classical estimators of μ_y are compared for a particular sampling scheme. A bivariate normal population with parameters $(\mu_y, \mu_x, \rho, \sigma_y^2, \sigma_x^2)$ is assumed. A bivariate sample of size n and an additional sample of size n_x on $\tilde{x}$ alone constitute the available sample data which are obtained from a one stage sample.

B. The Three Classical Estimators when the Covariance Matrix of $(\tilde{x}, \tilde{y})$ is Known

1. The estimator PT

The estimator PT is obtained by the estimation scheme given in Figure 6.1 with $\beta = 0$. The test statistic $z = \bar{y}_n - \bar{x}_{n+n_x}$ is used to make the preliminary test. If H_o: $\mu_y = \mu_x$ is accepted as a result of the preliminary test, then the estimator $(w_1\bar{y}_n + w_2\bar{x}_{n+n_x})$ is used. If H_A: $\mu_y \neq \mu_x$ is accepted, then the estimator $\bar{y}_n$ is used. The weights w_1 and w_2, given by equations (6.7.11) and (6.5.5), minimize the pooled estimator variance.

Using equation (6.5.12) with $\beta = 0$ and $w_2 = k_1/(k_1+k_2)$ yields the bias of the PT estimator as

$$B(PT) = -\frac{k_1}{\sigma_z}\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}(t+\delta)\,\phi(t)\,dt \qquad (8.2.1)$$

where

$$k_1 = \sigma_y^2/n - \rho\sigma_x\sigma_y/(n+n_x)$$

$$k_2 = \sigma_x^2/(n+n_x) - \rho\sigma_x\sigma_y/(n+n_x)$$

$$\sigma_z^2 = k_1 + k_2 \tag{8.2.2}$$

$$\delta = \Delta/\sigma_z$$

$$\Delta = \mu_y - \mu_x$$

Likewise, using equations (6.6.8) through (6.6.11) with $\beta = 0$ and $w_2 = k_1/(k_1 + k_2)$ yields the mean square error of PT as

$$\mathrm{MSE(PT)} = \frac{\sigma_y^2}{n} + \frac{k_1^2}{\sigma_z^2} \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta} (\delta^2-t^2)\ \phi(t)\ dt\,, \tag{8.2.3}$$

2. <u>The estimator PTR</u>

The estimator PTR is the regression estimator discussed in Chapter VI. Thus, if H_o: $\mu_y = \mu_x$ is accepted, PTR is the pooled estimator $(w_1\bar{y}_n + w_2\bar{x}_{n+n_x})$, where $w_2 = k_1/(k_1+k_2)$ as in the estimator PT. If H_A: $\mu_y \neq \mu_x$ is accepted, then PTR is the regression estimator $\left[\bar{y}_n + \beta(\bar{x}_{n+n_x} - \bar{x}_n)\right]$, where $\beta = \rho\sigma_y/\sigma_x$.

The bias of PTR is given by equation (6.9.1), i.e.

$$B(PTR) = -\frac{1}{\sigma_z}\int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left[k_1\delta + t\left\{k_1 - \frac{\rho^2\sigma_y^2 n_x}{n(n+n_x)}\right\}\right]\phi(t)\ dt. \qquad (8.2.4)$$

The mean square error of PTR is given by equation (6.9.2), i.e.

$$MSE(PTR) = \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{(n+n_x)}\right\} + \int_{-\xi_\alpha-\delta}^{\xi_\alpha-\delta}\left[K_2 t^2 + K_0\right]\phi(t)\ dt \qquad (8.2.5)$$

where

$$K_2 = -\frac{1}{\sigma_z^2}\left\{k_1 - \frac{\rho^2\sigma_y^2 n_x}{n(n+n_x)}\right\}^2 \qquad (8.2.6)$$

and

$$K_0 = \frac{1}{\sigma_z^2}\left\{k_1^2\delta^2 + \frac{\rho^2\sigma_y^2 n_x(k_2-k_1)}{n(n+n_x)} + \frac{\rho^4\sigma_y^4 n_x^2}{n^2(n+n_x)^2}\right\}.$$

3. The estimator R

An alternative to the estimators PT and PTR, both of which involve preliminary tests, is to do no preliminary test and always use a regression estimator. The regression estimator R is thus defined as

$$R = \bar{y}_n + \rho\sigma_y(\bar{x}_{n+n_x} - \bar{x}_n)/\sigma_x. \qquad (8.2.7)$$

The bias, variance, and mean square error of R are easily found to be

$$B(R) = 0 \tag{8.2.8}$$

and

$$V(R) = MSE(R) = \frac{\sigma_y^2}{n}\left\{1 - \frac{\rho^2 n_x}{n(n+n_x)}\right\} . \tag{8.2.9}$$

Note that $\bar{x}_{n+n_x}$ is in the position usually assumed by μ_x in the regression estimator. R is analogous to a regression estimator obtained from double sampling.

4. Special case when $\rho = 0$

If it is known that $\rho = 0$, i.e. $\tilde{x}$ and $\tilde{y}$ are independently normally distributed, then the estimator PTR reduces to the estimator PT, and the estimator R reduces to $\bar{y}_n$. Thus, the only remaining relevant comparison within the classical estimators is the never pool estimator $\bar{y}_n$ versus the sometimes pool estimator PT.

C. The Two Bayesian Estimators when the Covariance Matrix of $(\tilde{x}, \tilde{y})$ is Known

1. The prior distributions

To compare the Bayesian approach and the classical approach to pooling correlated data (in the sampling framework of section A of this chapter), only two of the different Bayesian estimators presented in Chapter VII are used here. These two estimators assume a $N(0, a^2)$ prior on $\tilde{\Delta}$. The willingness of an experimenter to use a preliminary test procedure or to put a

$N(0, a^2)$ prior distribution on $\tilde{\Delta}$ indicates that there exists a certain amount of belief that $\mu_y = \mu_x$. The preliminary test approach and the Bayesian approach of putting a $N(0, a^2)$ prior on $\tilde{\Delta}$ are then just two approaches to the same pooling problem.

2. Basis of comparison

There is no obvious basis of comparison for Bayesian and classical estimators. One criterion is to average the quantity $(\mu_y - \hat{\mu}_y)^2$ over the posterior distribution of $\tilde{\mu}_y$, and then to average the result over the unconditional distribution of the sample. This is equivalent to averaging $(\mu_y - \hat{\mu}_y)^2$ over the conditional distribution of the sample, given μ_y (which is the mean square error), and then averaging the result over the prior distribution of $\tilde{\mu}_y$. This comparison, however, favors the Bayesian estimator since the Bayesian estimator is obtained so as to minimize the posterior loss. Another possible criterion is to average $(\mu_y - \hat{\mu}_y)^2$ over the posterior distribution of $\tilde{\mu}_y$. However, if the Bayesian takes as his prior distribution the density which assigns a probability of one to some constant, then the above criterion is always zero for the Bayesian. A third criterion is to average $(\mu_y - \hat{\mu}_y)^2$ over the distribution of the sample, given μ_y, which is just mean square error.

Roberts (1966) discusses some of the problems inherent in such an investigation, and gives some suggestions for a criterion of comparison. First, he suggests that the model which is generating the data be known to both the Bayesian and the classicist. In this discussion, the model is a bivariate normal distribution. Second, he suggests a criterion which penalizes "an estimation error independently of the method that led to the error." (Roberts, 1966, p. 25). The mean square error is an example of such a criterion. Third, he discusses what values of the unknown parameter should be used in the comparison, i.e. whether one value is sufficient or whether several mean square errors for different values of the parameter ought to be averaged in some manner over the different parameter values. Roberts assumes throughout his discussion that the Bayesian has no prior information, i.e. the Bayesian is using a diffuse prior.

In this chapter the basis of comparison is mean square error. For the Bayesian estimators, mean square error is defined as $E(\tilde{\mu}_y - \hat{\mu}_y)^2$, where $\hat{\mu}_y$ is the Bayesian estimator and the expectation is taken with respect to the distribution of the sample data, given the parameters μ_y and Δ. Thus, the mean square error for the Bayesian estimators will be a function of the population parameters μ_y and Δ and the specified parameters of the prior distributions.

3. The estimator BN

In this chapter, since $n_y = 0$, then Q_y, Q_x, and Q_Δ as defined in equations (7.5.1) and (7.9.6) reduce to the following:

$$Q_x = \bar{x}_{n+n_x}$$

$$Q_y = \bar{y}_n + \rho\sigma_y(\bar{x}_{n+n_x} - \bar{x}_n)/\sigma_x \qquad (8.3.1)$$

$$Q_\Delta = \frac{n(\bar{y}_n - \bar{x}_n) + n_x\left[\bar{y}_n + \rho\sigma_y(\bar{x}_{n_x} - \bar{x}_n)/\sigma_x - \bar{x}_{n_x}\right]}{(n + n_x)} \quad .$$

Note that Q_y reduces to the regression estimator defined in the previous section as R. Recall also that Q_y, including the data $\bar{y}_{n_y}$, was the Bayesian estimator obtained in section E of Chapter VII when the principle of precise measurement was applied to $(\tilde{\mu}_y, \tilde{\mu}_x)$. Hence, the classical estimator R can also be considered as a Bayesian estimator when the bivariate prior on $(\tilde{\mu}_y, \tilde{\mu}_x)$ or $(\tilde{\mu}_y, \tilde{\Delta})$ is diffuse. R is the Bayesian estimator which is obtained under minimal specification of the prior distribution, and thus any further specification of the prior distribution should lead to an improved estimator.

From equations (7.5.2) through (7.5.5) and (7.9.9), with $n_y = 0$, the covariance elements of $(\tilde{Q}_x, \tilde{Q}_y)$, given (μ_x, μ_y), and $(\tilde{Q}_y, \tilde{Q}_\Delta)$, given (μ_y, Δ), for this chapter are

$$S_y^2 = \sigma_y^2/\left[M_y(1-R_{xy}^2)\right]$$

$$S_x^2 = \sigma_x^2/\left[M_x(1-R_{xy}^2)\right]$$

$$S_{xy}^2 = R_{xy}S_yS_x$$

$$R_{xy} = \rho\sqrt{n}\,/\left[\sqrt{1-\rho^2}\,\sqrt{M_x}\,\right] \qquad (8.3.2)$$

$$M_y = n/(1-\rho^2)$$

$$M_x = n_x + n/(1-\rho^2)$$

$$S_\Delta^2 = S_y^2 + S_x^2 - 2S_{xy}^{\ 2}$$

$$S_{y\Delta}^{\ 2} = S_y^2 - S_{xy}^{\ 2} = R_{y\Delta}S_yS_\Delta$$

The estimator BN is the Bayesian estimator discussed in section H of Chapter VII. It is obtained by assuming $\tilde{\mu}_y$ and $\tilde{\Delta}$ to be independently normally distributed, i.e. $\tilde{\mu}_y$ is $N(b_y, \alpha_y^2)$ and $\tilde{\Delta}$ is $N(0, a^2)$. From equation (7.8.5), the estimator BN is obtained as

$$BN = \frac{1}{D}\left[\begin{array}{l} b_y\left\{\dfrac{1}{\alpha_y^2 a^2} - \dfrac{R_{xy}}{a^2 S_x S_y(1-R_{xy}^{\ 2})} + \dfrac{1}{S_x^2\alpha_y^2(1-R_{xy}^{\ 2})}\right\} \\ + \dfrac{Q_y}{S_xS_y(1-R_{xy}^{\ 2})}\left\{\dfrac{1}{S_xS_y} - \dfrac{R_{xy}}{a^2}\right\} + \dfrac{Q_x}{a^2S_x^2(1-R_{xy}^{\ 2})} \\ + \dfrac{\left[Q_y + R_{xy}S_y(b_y - Q_x)/S_x\right]}{a^2S_y^2(1-R_{xy}^{\ 2})} \end{array}\right] \qquad (8.3.3)$$

where D is obtained from equation (7.8.4) as

$$D = \left\{\frac{1}{a^2} + \frac{1}{S_x^2(1-R_{xy}^{\ 2})}\right\}\left\{\frac{1}{a^2} + \frac{1}{\alpha_y^2} + \frac{1}{S_y^2(1-R_{xy}^{\ 2})}\right\} - \left\{\frac{1}{a^2} + \frac{R_{xy}}{S_xS_y(1-R_{xy}^{\ 2})}\right\}^2 . \qquad (8.3.4)$$

The expectation of BN is defined to be the expectation over the conditional distribution of $(\tilde{Q}_y, \tilde{Q}_x)$, given (μ_y, Δ). The bias of BN is then easily obtained as

$$B(BN) = \frac{1}{D}\left[\frac{(b_y-\mu_y)}{\alpha_y^2}\left\{\frac{1}{a^2}+\frac{1}{S_x^2(1-R_{xy}^2)}\right\} - \frac{\Delta}{a^2(1-R_{xy}^2)}\left\{\frac{1}{S_x^2}-\frac{R_{xy}}{S_xS_y}\right\}\right] \tag{8.3.5}$$

where D is given in equation (8.3.4). Bias is not a relevant concept to Bayesians since their only interest is in minimizing the posterior loss. However, bias is calculated here as a step toward obtaining the mean square error criterion for the Bayesisn estimators. In equation (8.3.5) note that the bias of BN depends upon $(b_y - \mu_y)$ and Δ. In general, the bias increases as $|\Delta|$ and $|b_y - \mu_y|$ increase, although a cancellation effect could operate to produce a bias near zero for large $|\Delta|$ and $|b_y - \mu_y|$. Note also in equation (8.3.5) that the factor $(b_y - \mu_y)$ is weighted inversely proportional to α_y^2 and the quantity Δ is weighted inversely proportional to a^2. This occurs because b_y is not weighted heavily in the estimator BN if α_y^2 is large. Hence, even if b_y is not a good estimator of μ_y, it will not contribute very much to the bias. A similar statement holds for Δ and a^2.

After some algebra, the variance of the estimator BN is obtained as

$$V(BN) = \frac{1}{D^2}\left[\frac{1}{S_x^4S_y^2(1-R_{xy}^2)^2}+\frac{1}{a^2(1-R_{xy}^2)}\left\{\frac{1}{a^2S_x^2}-\frac{2R_{xy}}{a^2S_xS_y}+\frac{1}{a^2S_y^2}+\frac{2}{S_x^2S_y^2}\right\}\right] \tag{8.3.6}$$

where D is given in equation (8.3.4). The variance thus depends upon the quantities S_x^2, S_y^2, R_{xy}, and the parameters a^2 and α_y^2 of the prior distributions. The mean square error of BN, i.e. MSE(BN), is obtained as

$$MSE(BN) = V(BN) + \left[B(BN)\right]^2. \tag{8.3.7}$$

If $a^2 \rightarrow \infty$, it is easily shown that

$$\lim_{a^2\rightarrow\infty} BN = \left[b_y/\alpha_y^2 + Q_y/s_y^2\right]\left[1/\alpha_y^2 + 1/s_y^2\right]^{-1}, \tag{8.3.8}$$

$$\lim_{a^2\rightarrow\infty} B(BN) = \frac{(b_y - \mu_y)}{\alpha_y^2}\left[\frac{1}{s_y^2} + \frac{1}{\alpha_y^2}\right]^{-1} \tag{8.3.9}$$

$$\lim_{a^2\rightarrow\infty} V(BN) = \left[1/\alpha_y^2 + 1/s_y^2\right]^{-2}/s_y^2. \tag{8.3.10}$$

In the limit, the estimator BN does not utilize Q_x and hence the bias is not a function of $\Delta = \mu_y - \mu_x$. The estimator is a weighted average of b_y and Q_y, and since $E(\tilde{Q}_y \mid \mu_y, \Delta) = \mu_y$, then the magnitude of the bias depends upon $|b_y - \mu_y|$.

If now $\alpha_y^2 \rightarrow \infty$, then a little algebra shows that

$$\lim_{\alpha_y^2\rightarrow\infty} BN = \frac{\left[\frac{Q_x}{a^2}\left\{\frac{1}{s_x^2} - \frac{R_{xy}}{s_x s_y}\right\} + Q_y\left\{\frac{1}{s_x^2 s_y^2} + \frac{1}{a^2 s_y^2} - \frac{R_{xy}}{a^2 s_x s_y}\right\}\right]}{\left[\frac{1}{a^2}\left\{\frac{1}{s_y^2} + \frac{1}{s_x^2} - \frac{2R_{xy}}{s_x s_y}\right\} + \frac{1}{s_x^2 s_y^2}\right]}, \tag{8.3.11}$$

$$\lim_{\alpha_y^2 \to \infty} B(BN) = \frac{-\dfrac{\Delta}{a^2}\left\{\dfrac{1}{S_x^2} - \dfrac{R_{xy}}{S_x S_y}\right\}}{\left[\dfrac{1}{a^2}\left\{\dfrac{1}{S_x^2} + \dfrac{1}{S_y^2} - \dfrac{2R_{xy}}{S_x S_y}\right\} + \dfrac{1}{S_x^2 S_y^2}\right]} \tag{8.3.12}$$

and

$$\lim_{\alpha_y^2 \to \infty} V(BN) = \frac{(1-R_{xy}^2)\left[\dfrac{1}{a^4}\left\{\dfrac{1}{S_x^2} + \dfrac{1}{S_y^2} - \dfrac{2R_{xy}}{S_x S_y}\right\} + \dfrac{2}{a^2 S_x^2 S_y^2} + \dfrac{1}{S_x^4 S_y^2 (1-R_{xy}^2)}\right]}{\left[\dfrac{1}{a^2}\left\{\dfrac{1}{S_y^2} + \dfrac{1}{S_x^2} - \dfrac{2R_{xy}}{S_x S_y}\right\} + \dfrac{1}{S_x^2 S_y^2}\right]^2} . \tag{8.3.13}$$

Note that the estimator in equation (8.3.11) is a weighted average of Q_y and Q_x, where the weight on Q_x is inversely proportional to a^2. Also note that the magnitude of the bias depends upon $|\Delta|$.

If now $\alpha_y^2 \longrightarrow \infty$ and $a^2 \longrightarrow \infty$, then

$$\lim_{\substack{\alpha_y^2 \to \infty \\ a^2 \to \infty}} BN = Q_y = R \tag{8.3.14}$$

$$\lim_{\substack{\alpha_y^2 \to \infty \\ a^2 \to \infty}} B(BN) = 0 \tag{8.3.15}$$

$$\lim_{\substack{\alpha_y^2 \to \infty \\ a^2 \to \infty}} V(BN) = S_y^2. \tag{8.3.16}$$

The limiting estimator in equation (8.3.14) is also the estimator R as in equation (8.2.7) and is, in addition, the Bayesian estimator when the principle of precise measurement is applied to the joint distribution of $(\tilde{\mu}_y, \tilde{\mu}_x)$ or $(\tilde{\mu}_y, \tilde{\Delta})$.

If $a^2 \longrightarrow 0$, then it is easy to show that

$$\lim_{a^2 \to 0} BN = \frac{\left[\frac{b_y}{\alpha_y^2} + \frac{Q_y}{(1-R_{xy}^2)}\left\{\frac{1}{S_y^2} - \frac{R_{xy}}{S_x S_y}\right\} + \frac{Q_x}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} - \frac{R_{xy}}{S_x S_y}\right\}\right]}{\left[\frac{1}{\alpha_y^2} + \frac{1}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} + \frac{1}{S_y^2} - \frac{2R_{xy}}{S_x S_y}\right\}\right]} \tag{8.3.17}$$

$$\lim_{a^2 \to 0} B(BN) = \frac{\left[\frac{(b_y - \mu_y)}{\alpha_y^2} - \frac{\Delta}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} - \frac{R_{xy}}{S_x S_y}\right\}\right]}{\left[\frac{1}{\alpha_y^2} + \frac{1}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} + \frac{1}{S_y^2} - \frac{2R_{xy}}{S_x S_y}\right\}\right]} \tag{8.3.18}$$

and

$$\lim_{a^2 \to 0} V(BN) = \frac{\frac{1}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} + \frac{1}{S_y^2} - \frac{2R_{xy}}{S_x S_y}\right\}}{\left[\frac{1}{\alpha_y^2} + \frac{1}{(1-R_{xy}^2)}\left\{\frac{1}{S_x^2} + \frac{1}{S_y^2} - \frac{2R_{xy}}{S_x S_y}\right\}\right]^2}. \tag{8.3.19}$$

In this case, Q_y and Q_x are considered equally reliable estimators of μ_y and are weighted so as to minimize the "variance" of the pooled estimator $(w_1 b_y + w_2 Q_y + [1 - w_1 - w_2] Q_x)$, where the term "variance" has been defined in section H of Chapter VII. Note that the bias of BN in this limiting case depends both upon Δ and $(b_y - \mu_y)$. Even though $a^2 \longrightarrow 0$, which is equivalent to the prior probability statement $Pr[\tilde{\Delta} = 0] = 1$, the bias still depends upon $\Delta = \mu_y - \mu_x$ since the parameter Δ may, in fact, be different from zero.

If, now, $\alpha_y^2 \longrightarrow 0$, then it is easy to show that

$$\lim_{\alpha_y^2 \to 0} BN = b_y, \tag{8.3.20}$$

$$\lim_{\alpha_y^2 \to 0} B(BN) = (b_y - \mu_y), \tag{8.3.21}$$

and

$$\lim_{\alpha_y^2 \to 0} V(BN) = 0. \tag{8.3.22}$$

In this case the estimator of μ_y is always b_y and is independent of the sample results. This occurs, of course, because in the limiting case as $\alpha_y^2 \longrightarrow 0$, the prior distribution on $\tilde{\mu}_y$ is degenerate and has a probability of one at the point b_y. In such a case, the posterior distribution of $\tilde{\mu}_y$ must always equal the prior distribution.

Consider now the limit of the estimator BN as $a^2 \longrightarrow 0$ and $\alpha_y^2 \longrightarrow \infty$. Then it is easily shown that

$$\lim_{\substack{a^2 \to 0 \\ \alpha_y^2 \to \infty}} BN = \frac{\left[Q_y \left\{\frac{1}{s_y^2} - \frac{R_{xy}}{s_x s_y}\right\} + Q_x \left\{\frac{1}{s_x^2} - \frac{R_{xy}}{s_x s_y}\right\}\right]}{\left\{\frac{1}{s_x^2} + \frac{1}{s_y^2} - \frac{2R_{xy}}{s_x s_y}\right\}} \tag{8.3.23}$$

$$\lim_{\substack{a^2 \to 0 \\ \alpha_y^2 \to \infty}} B(BN) = \frac{-\Delta \left\{\frac{1}{s_x^2} - \frac{R_{xy}}{s_x s_y}\right\}}{\left\{\frac{1}{s_x^2} + \frac{1}{s_y^2} - \frac{2R_{xy}}{s_x s_y}\right\}}, \tag{8.3.24}$$

and

$$\lim_{\substack{a^2 \to 0 \\ \alpha_y^2 \to \infty}} V(BN) = (1-R_{xy}^2) \left\{\frac{1}{s_x^2} + \frac{1}{s_y^2} - \frac{2R_{xy}}{s_x s_y}\right\}^{-1} . \tag{8.3.25}$$

Hence, the estimator is a weighted average of Q_y and Q_x, where the weights minimize $V\left[w_1\tilde{Q}_y + (1 - w_1)\tilde{Q}_x\right]$. Note that the magnitude of the bias depends upon $|\Delta|$.

4. The estimator BPM

The estimator BPM is the estimator discussed in section I of Chapter VII. It is obtained by assuming $\tilde{\Delta}$ and $\tilde{\mu}_y$ to be independently distributed, where $\tilde{\Delta}$ is $N(0,a^2)$ and the principle of precise measurement is applied to $\tilde{\mu}_y$. BPM is given by Z_y in equation (7.9.14), but it was shown in section I of Chapter VII that Z_y could also be obtained as the limit when $\alpha_y^2 \to \infty$ of the posterior mean when $\tilde{\Delta}$ is $N(0,a^2)$ and $\tilde{\mu}_y$ is $N(b_y, \alpha_y^2)$. Thus, BPM can be obtained as the limit of BN as $\alpha_y^2 \longrightarrow \infty$. Equations (8.3.11), (8.3.12), and (8.3.13) thus give the estimator BPM, its bias B(BPM), and variance V(BPM).

Note that both of the estimators BN and BPM utilize regression estimators. BPM utilizes regression because Q_y is a regression of $\bar{y}_n$ upon $\bar{x}_n$. BN, in addition, also uses a regression of Q_y upon Q_x. Of the three classical estimators mentioned in the preceding section, R and PTR use regression, whereas PT does not.

5. Special case when $\rho = 0$

If $\rho = 0$, i.e. $\tilde{x}$ and $\tilde{y}$ are independently, normally distributed, then the statistics Q_y, Q_x, and Q_Δ in this chapter reduce to

$$\begin{aligned} Q_x &= \bar{x}_{n+n_x} \\ Q_y &= \bar{y}_n \qquad\qquad \rho = 0,\ n > 0,\ n_x > 0. \\ Q_\Delta &= \bar{y}_n - \bar{x}_{n+n_x} \end{aligned} \tag{8.3.26}$$

For $\rho = 0$, and $n > 0$, $n_x > 0$, the estimator BN reduces to

$$BN = \frac{\left[b_y/\alpha_y^2 + \bar{x}_{n+n_x} / \left[a^2+\sigma_x^2/(n+n_x)\right] + n\bar{y}_n/\sigma_y^2 \right]}{\left[1/\alpha_y^2 + 1/\left[a^2+\sigma_x^2/(n+n_x)\right] + n/\sigma_y^2 \right]} \tag{8.3.27}$$

The bias and variance for this special case are easily obtained from equations (8.3.5) and (8.3.6). Note that the estimator in equation (8.3.27) is a weighted average of the three "independent" components Q_x, Q_y, and b_y, where each component is weighted inversely proportional to its "variance". The bias depends upon both $(b_y - \mu_y)$ and Δ.

The estimator BPM, when $\rho = 0$, can be obtained from equation (8.3.27) by letting $\sigma_y^2 \longrightarrow \infty$. Hence,

$$\text{BPM} = \frac{\left[\bar{x}_{n+n_x}/\left[a^2 + \sigma_x^2/(n+n_x)\right] + n\bar{y}_n/\sigma_y^2\right]}{\left[1/\left[a^2 + \sigma_x^2/(n+n_x)\right] + n/\sigma_y^2\right]}, \ \rho = 0 \qquad (8.3.28)$$

The bias and variance for this special case can be obtained from equations (8.3.5) and (8.3.6). The estimator in equation (8.3.28) is close to the classical preliminary test pooled estimator if a^2 is small and, likewise, close to the classical preliminary test non-pooled estimator if a^2 is large.

Neither of the estimators BN nor BPM contains a regression term when $\rho = 0$, and further, Q_y does not contain the regression component as in equation (8.3.1) for $\rho \neq 0$. Both estimators are simply weighted averages of $\bar{y}_n$ and $\bar{x}_{n+n_x}$, where the weight on $\bar{x}_{n+n_x}$ decreases as a^2 increases.

D. The Three Classical Estimators when the Covariance Matrix is Unknown

1. Theoretical difficulties and Monte Carlo procedure

Some difficulties of the preliminary test approach under the assumption of unknown covariance matrix were mentioned in section C of Chapter III. The main point there was that, for $\rho = 0$, the assumption that $\sigma_y^2 \neq \sigma_x^2$ leads naturally to the test statistic $(\bar{y}_n - \bar{x}_{n+n_x})/\sqrt{s_y^2/n + s_x^2/(n+n_x)}$, where s_y^2

and s_x^2 are the sample variances. This test statistic does not follow the t-distribution, of course, but it does follow the Behrens-Fisher distribution. Recall, however, that in deriving the bias and mean square error of a preliminary test type estimator, the joint distribution of the test statistic and the possible estimators is required. Thus, a derivation of bias and mean square error of a preliminary test type estimator for $\rho = 0$ and unknown $\sigma_x^2 \neq \sigma_y^2$ would require evaluating probabilities of a joint distribution of two or three dependent random variables, one of which follows the Behrens-Fisher distribution. This looks rather intractable.

If $\rho \neq 0$, there is no common test statistic to use for the preliminary test. A tractable test statistic is

$$t = \sqrt{n}\,(\bar{y}_n - \bar{x}_n) \Big/ \sqrt{s_{xn}^2 + s_{yn}^2 - 2s_{xyn}^2}\ , \qquad (8.4.1)$$

where s_{xn}^2, s_{yn}^2, and s_{xyn}^2 are the usual unbiased estimates of σ_x^2, σ_y^2, and σ_{xy}^2, respectively, from the bivariate sample of size n. The test statistic t in equation (8.4.1) follows the Student-t distribution with (n-1) df, but does not incorporate all of the available sample data since n_x extra observations on $\tilde{x}$ are not used. If now the estimators PT and PTR of section B have extended definitions for the case of unknown covariance matrix (i.e. estimate β and w_2 from the sample), then the bias and mean square error of these new estimators can be obtained by considering the joint distribution of t in equation (8.4.1) and the possible estimators. Writing down the first few lines of the expectation of $\hat{\mu}_y$ shows that the joint distribution of $(\beta, \bar{x}_n, t)$ and $(\hat{w}_2, \bar{y}_n, t)$, as well as some others, are needed. This problem is certainly not trivial, but some progress can probably be made on it. However, the solution to the problem may not be sufficiently general to warrant the complicated derivation.

If it is desired to use all of the data to make the preliminary test, then a possible test statistic is t^*, where

$$t^* = (\bar{y}_n - \bar{x}_{n+n_x}) \Big/ \sqrt{\frac{s^2_{yn}}{n} + \frac{s^2_{x(n+n_x)}}{(n+n_x)} - \frac{2s^2_{xyn}}{(n+n_x)}} \tag{8.4.2}$$

and $s^2_{x(n+n_x)}$ is the usual unbiased sample estimate of σ^2_x which is based on $(n+n_x)$ observations. It is not at all clear, however, what distribution t^* follows. Hence, a derivation of the bias and mean square error looks rather intractable for the estimators PT and PTR under unknown covariance matrix when the test statistic t^* is used to make the preliminary test.

In an effort to discover, however, some properties of the preliminary test type estimators when the covariance matrix is not known, the three classical estimators discussed in section B of this chapter are defined below for the case of unknown covariance matrix. Their bias, variance, and mean square error are obtained by Monte Carlo methods, the results of which are reported in section F of this chapter.

The three estimators defined in this section are called PTU, PTRU, and RU. For the purpose of the Monte Carlo study, a bivariate normal population with parameters $(\sigma^2_y, \sigma^2_x) = (6,12)$ was chosen. The bivariate sample size was $n = 9$, and the additional sample on $\tilde{x}$ was of size $n_x = 18$. Also considered were $n = 25$ and $n_x = 20$. Values of $\rho = 1/3$ and $\rho = 0$ were investigated for values of Δ ranging from 0 to 4.0 and $\alpha = .50, .25, .10$. μ_y was always taken to be zero.

For each specification of $(\Delta, \sigma_y^2, \sigma_x^2, \rho, n_x, n)$, then, the following prodedure was repeated one thousand times: (1) select a random sample of size n from the bivariate normal distribution with parameters ($\mu_y = 0$, $\mu_x = -\Delta$, σ_y^2, σ_x^2, ρ), and select a random sample of size n_x from $N(-\Delta, \sigma_x^2)$; (2) perform a preliminary test of H_o: $\mu_y = \mu_x$ versus H_A: $\mu_y \neq \mu_x$ using only the sample data; (3) on the basis of the outcome of the preliminary test in step (2), compute the estimators PTU and PTRU.

After the above procedure was repeated one thousand times, the means, variances, and mean square errors of the thousand estimates were computed for PTU and PTRU. These values serve as estimates of the true bias and mean square error of the estimators PTU and PTRU.

The preliminary test of step (2) was done in two different ways, the results of which are denoted in section F of this chapter by Monte Carlo I and Monte Carlo II. In Monte Carlo I the preliminary test of H_o: $\mu_y = \mu_x$ versus H_A: $\mu_y \neq \mu_x$ was done using the test statistic t as given in equation (8.4.1), which was compared to a critical value from the student-t distribution for the appropriate α and n. In Monte Carlo II, the test statistic t^*, as given in equation (8.4.2), was used for the preliminary test. This test statistic is also obtained by following the preliminary test procedure as outlined in Chapter III with $m_y = m_x = n_y = 0$ and substituting sample estimates for unknown parameter values. Since t^* does not follow a Student-t distribution, an approximate critical value for t^* for a given α level was obtained by using the corresponding critical value from the N(0,1) distribution for the given α level.

2. The estimator PTU

The estimator PTU is obtained from the estimator PT discussed in section B of this chapter, by replacing the elements of the unknown covariance matrix by their sample estimates. The quantities k_1 and k_2 of equation (8.2.2) are estimated from the sample by

$$\begin{aligned} \hat{k}_1 &= s^2_{yn}/n - s^2_{xyn}/(n+n_x) \\ \hat{k}_2 &= s^2_{x(n+n_x)}/(n+n_x) - s^2_{xyn}/(n+n_x) \end{aligned} \tag{8.4.3}$$

and hence the weights w_1 and w_2 are estimated from the sample by

$$\begin{aligned} \hat{w}_1 &= \hat{k}_2/(\hat{k}_1 + \hat{k}_2) \\ \hat{w}_2 &= 1 - \hat{w}_1. \end{aligned} \tag{8.4.4}$$

The estimator PTU is then defined as

$$\text{PTU} = \begin{Bmatrix} \hat{w}_1\,\bar{y}_n + \hat{w}_2\,\bar{x}_{n+n_x} & \text{if } H_o\text{: } \mu_y = \mu_x \text{ accepted} \\ \bar{y}_n & \text{if } H_A\text{: } \mu_y \neq \mu_x \text{ accepted} \end{Bmatrix}. \tag{8.4.5}$$

Recall that the preliminary test of H_o versus H_A was done two different ways, i.e. by Monte Carlo I and Monte Carlo II. These two procedures most likely have a different Type I error. However, in both procedures the same non-pooled estimator $\bar{y}_n$ and the same pooled estimator $\left\{w_1 \bar{y}_n + w_2 \bar{x}_{n+n_x}\right\}$ is used.

3. The estimator PTRU

The estimator PTRU is obtained from the estimator PTR discussed in section B of this chapter by replacing unknown parameters by their sample estimates. An estimate of the regression coefficient from the sample is

$$\hat{\beta} = s^2_{xyn}/s^2_{xn} \; . \tag{8.4.6}$$

The estimator PTRU is then defined as

$$\text{PTRU} = \begin{bmatrix} \hat{w}_1\bar{y}_n + \hat{w}_2\bar{x}_{n+n_x} & \text{if } H_o\text{: } \mu_y = \mu_x \text{ accepted} \\ & \\ \bar{y}_n + \hat{\beta}(\bar{x}_{n+n_x} - \bar{x}_n) & \text{if } H_A\text{: } \mu_y \neq \mu_x \text{ accepted} \end{bmatrix} \tag{8.4.7}$$

where $\hat{w}_1$ and $\hat{w}_2$ are the same weights used in PTU and defined in equation (8.4.4). Again, the preliminary test was done two ways, using t as in equation (8.4.1) and t^* as in equation (8.4.2).

4. The estimator RU

The estimator RU is obtained from the estimator R in section B of this chapter by replacing β with the sample estimate $\hat{\beta}$. No preliminary test is done in order to obtain RU, and the estimator is simply

$$RU = \bar{y}_n + \hat{\beta}(\bar{x}_{n+n_x} - \bar{x}_n). \tag{8.4.8}$$

Hence, RU is the same stimator as PTRU when $H_A: \mu_y \neq \mu_x$ is accepted.

Since the estimator RU involves no preliminary test, its bias, variance, and mean square error are more tractable. Note that the regression estimator in equation (8.4.8) is not the usual one where the x_i's are assumed to be fixed constants. In equation (8.4.8), both $\tilde{x}_i$ and $\tilde{y}_i$ are random variables, and their joint distribution is bivariate normal. Tikkiwal (1960) shows that $\bar{x}_n$ and $\bar{y}_n$ are statistically independent of $\hat{\beta}$, and hence the estimator RU is unbiased. He further derives the variance of RU as

$$V(RU) = \frac{n_x \sigma_y^2 (1-\rho^2)}{n(n+n_x)} \left\{ 1 + \frac{1}{(n-3)} \right\} + \frac{\sigma_y^2}{(n+n_x)} \tag{8.4.9}$$

and also gives an unbiased estimator of V(RU).

In the Monte Carlo study reported in section F of this chapter, the Monte Carlo estimates of bias and mean square error of RU were obtained since only a few very minor additional calculations were needed. The true variance as given in equation (8.4.9) was then compared to the Monte Carlo bias and variance to give some indication about the precision of the Monte Carlo results.

5. Special case when $\rho = 0$

If it is known that $\rho = 0$, where $\sigma_y^2 \neq \sigma_x^2$ are still unknown, then the estimator PTRU is the same as PTU, and the estimator RU reduces to $\bar{y}_n$.

Thus, the only remaining comparison is the never pool estimator $\bar{y}_n$ compared to the sometimes pool estimator PRU. If it is known that $\rho = 0$, then the test statistic for the preliminary test is defined as

$$t' = (\bar{y}_n - \bar{x}_{n+n_x}) \Big/ \sqrt{s^2_{yn}/n + s^2_{x(n+n_x)}/(n+n_x)} \quad . \tag{8.4.10}$$

In Monte Carlo I, the test statistic t' was compared to a critical value obtained by using the Satterthwaite approximation. In Monte Carlo II, t' was compared to the critical value from the N(0,1) distribution for the given α level of the preliminary test.

E. The Bayesian Estimators when the Covariance Matrix of $(\tilde{x}, \tilde{y})$ is Unknown

1. General approach and literature review

The Bayesian approach to estimating μ_y when the covariance matrix of $(\tilde{x}, \tilde{y})$ is unknown consists of assigning some prior distribution to the elements of the covariance matrix as well as the prior distribution on the population means. Then, the posterior distribution of $(\tilde{\mu}_y, \tilde{\Delta})$ is obtained, in effect, by "integrating out" the variables ρ, σ^2_x, σ^2_y. The posterior distribution of $\tilde{\mu}_y$ is then obtained as the marginal of the above distribution.

The relevant question, then, is what priors on the covariance matrix are feasible. Recall that the data distribution is bivariate normal and the prior distributions considered in Chapter VII were either normal, constant (precise measurement), or uniform over a specified interval.

Raiffa and Schlaifer (1961) discuss the case of obtaining the posterior distribution of $(\tilde{\mu}_y, \tilde{\mu}_x)$ for a bivariate sample of size n from a bivariate normal population. However, they assume that the "relative precision", defined in Raiffa and Schlaifer (1961, p. 310), of the covariance matrix is known, i.e. the matrix

$$\begin{bmatrix} \dfrac{\sigma_y}{\sigma_x\sqrt{1-\rho^2}} & \dfrac{-\rho}{\sqrt{1-\rho^2}} \\ \dfrac{-\rho}{\sqrt{1-\rho^2}} & \dfrac{\sigma_x}{\sigma_y\sqrt{1-\rho^2}} \end{bmatrix} \tag{8.5.1}$$

is known. (If $\rho = 0$, this amounts to knowing the ratio σ_y/σ_x only, but not the particular values of σ_y and σ_x.) When the "relative precision" is known, the natural conjugate distribution is normal-gamma, where a bivariate normal prior distribution is taken on $(\tilde{\mu}_x, \tilde{\mu}_y)$ and a gamma prior distribution is taken on the scalar "mean precision" $|\Sigma^{-1}|^{\frac{1}{2}}$, where Σ is the unknown covariance matrix. The product of the "relative precision" matrix and the scalar "mean precision" is Σ^{-1}, the inverse of the covariance matrix. The posterior distribution of $(\tilde{\mu}_x, \tilde{\mu}_y, |\tilde{\Sigma}^{-1}|^{\frac{1}{2}})$ is also normal-gamma, and the posterior mean of $\tilde{\mu}_y$ is a weighted average of $\bar{x}_n$ and $\bar{y}_n$, where the weights are functions of the known matrix in equation (8.5.1). This presentation by Raiffa and Schlaifer (1961) also assumes that the prior covariance matrix of $(\tilde{\mu}_x, \tilde{\mu}_y)$ can be expressed in units of the unknown "mean precision" $|\Sigma^{-1}|^{\frac{1}{2}}$. All of the above assumptions seem too rigid for an unknown covariance matrix, and this solution is not pursued further. It should be noted that the formulas given by Raiffa and Schlaifer (1961) for the case of "relative precision" known reduce to the case of total covariance matrix known if ρ, σ_y, and σ_x are each assumed to be known rather than just assuming the ratios in equation (8.5.1) to be known.

Evans (1965) discusses the case noticeably lacking in Raiffa and Schlaifer (1961), i.e. the Bayesian estimation of the mean vector of a multivariate normal population when both the mean vector and covariance matrix are unknown. He defines the natural conjugate distribution so that the prior distribution of the vector mean $\underline{\tilde{\mu}}$, given the covariance matrix Σ of the data distribution, is multivariate normal with mean $\underline{\xi}$ and covariance matrix Σ/λ, where $\lambda > 0$ is a real number. Note that the prior covariance matrix on $\underline{\tilde{\mu}}$ is then simply a multiple of Σ. The prior distribution on $\tilde{\Sigma}$ is taken to be the Wishart distribution. Evans (1965) then shows that the optimum estimator of μ_y for a squared error loss function is

$$\hat{\underline{\mu}} = \frac{n\bar{\underline{x}} + \lambda\underline{\xi}}{(n+\lambda)}, \tag{8.5.2}$$

where $\bar{\underline{x}}$ is the vector of sample means from a multivariate sample of size n. Evans (1965, p. 282) says

"The simpler form...is due to the assumption, implicit in the

> natural conjugate joint prior density of $\underline{\mu}$ and Σ...that the variance of the conditional prior distribution on $\underline{\tilde{\mu}}$ is a scalar multiple of Σ."

Ando and Kaufman (1965) obtain the same result as Evans (1965) by using the natural conjugate normal-Wishart distribution.

Geisser (1965b) and Geisser and Cornfield (1963) consider diffuse priors on $\underline{\tilde{\mu}}$ and $\tilde{\Sigma}$ and obtain the sample mean vector as the estimator of the population mean vector. Since diffuse priors are used, it is hardly surprising that the estimator of μ_y in the bivariate normal case would then be just $\bar{y}_n$. Stone (1963, 1964) has some critical comments on the approach by Geisser and Cornfield (1963), the main point being that the diffuse prior on $\underline{\tilde{\mu}}$ does not integrate out to one over the real line. But, of course, that criticism can be directed toward many who work with diffuse priors.

From the previous comments, two conclusions can be drawn which are relevant to the problem of "pooling means" from a Bayesian viewpoint when the covariance matrix Σ is unknown. First, in the natural conjugate case above the estimator of μ_y is a weighted average of $\bar{y}_n$ and the prior mean of $\tilde{\mu}_y$. Hence, the estimator of μ_y incorporates none of the existing information on $\tilde{x}$. Second, inherent in the natural conjugate approach is the assumption that the prior covariance matrix of $(\tilde{\mu}_y, \tilde{\mu}_x)$ is a multiple of the data covariance matrix Σ. If two independent normal priors were desired on $\tilde{\mu}_y$ and $\tilde{\Delta}$, then this would imply that the data distribution have the parallel characteristic that $\tilde{y}$ and $(\tilde{y} - \tilde{x})$ are independent. This assumption is probably too restrictive for application. For these two reasons, the available literature does not seem to offer a solution to the Bayesian problem of "pooling means" with unknown covariance matrix.

2. The Bayesian estimators for the comparison study

Even if the approaches mentioned above evidenced some pooling of the x and y data in estimating μ_y, they would still have to be modified somewhat to include the sampling plan used in Chapters VII and/or VIII. Recall in Chapter VII that the estimator of μ_y was obtained by (1) proving that the statistics Q_x and Q_y were sufficient, (2) noting that Q_x and Q_y had a bivariate normal distribution, (3) applying the results of Raiffa and Schlaifer (1961) for bivariate sampling to $\tilde{Q}_x$ and $\tilde{Q}_y$, where the bivariate sample was of size one.

The derivation of the sufficient statistics for the unknown covariance matrix is not as simple as presented in Lemma 7.1 for the known covariance matrix. First of all, Q_x and Q_y are clearly no longer sufficient since they depend upon the parameters ρ, σ_y^2, and σ_x^2. A little work on the likelihood function shows that for $n > 0$, $n_x > 0$, $n_y > 0$, the statistics $(\bar{x}_n, \bar{y}_n, \bar{x}_{n_x}, \bar{y}_{n_y}, s_{xn}^2, s_{yn}^2, s_{xyn}^2, s_{xn_x}^2, s_{yn_y}^2)$ are sufficient, where $s_{yn_y}^2$ is the estimate of σ_y^2 from the sample of size n_y on $\tilde{y}$, with a similar definition for $s_{xn_x}^2$. This does not appear to be a minimal set, however. Of course, any prior can be taken on $(\tilde{\mu}_y, \tilde{\mu}_x, \tilde{\rho}, \tilde{\sigma}_x^2, \tilde{\sigma}_y^2)$, the above sufficient statistics used, and the posterior distribution of $(\tilde{\mu}_y, \tilde{\mu}_x)$ can be obtained.

Further investigation is needed on this problem, and hence no Bayesian estimators are compared in the next section under the assumption of unknown covariance matrix.

F. A Comparison of the Five Estimators

The following topics are investigated in this section: (1) the comparison of the estimators PT, PTR, and R under the assumption of known covariance matrix; (2) the comparison of the estimators PTU, PTRU, and RU under the assumption of unknown covariance matrix; (3) the comparison of PT, PTR, and R to PTU, PTRU, and RU to see the effect on bias and mean square error of substituting sample estimates for the unknown elements of the covariance matrix; (4) the comparison of the Bayesian estimators BN and BPM to the classical estimators PT, PTR, and R under the assumption of known covariance matrix. Comparisons (1) and (4) are made on the basis of theoretical results presented in sections B and C of this chapter; comparisons (2) and (3) are made on the basis of the Monte Carlo scheme discussed in section D of this chapter.

Tables 8.1 through 8.3 give the results of the comparison study of the classical estimators for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\rho = 1/3$, $(n,n_x)=(9,18)$, $\alpha=.50$, $.25$, $.10$, and $\Delta=0$, 1, 2. The values of δ corresponding to $\Delta=0$, 1, 2 are $\delta=0$, 1.05, 2.11, respectively.

From Table 8.1 where $\Delta=0$, MSE(PTR) is always less than MSE(R), which in turn is always less than MSE(PT). The Monte Carlo I estimate of MSE(PTU) is always less than the estimate of MSE(PTRU), but the two estimates are rather close for a given α level. Also, the estimates of MSE(PTRU), MSE(PTU), and MSE(RU) in most cases are not too different from the theoretical mean square error with known covariance matrix. The Monte Carlo estimate of V(RU) is 0.6584, whereas Tikkiwal (1960) gives 0.6831 as the theoretical value of V(RU). The estimates of MSE(PTRU) and MSE(PTU) from Monte Carlo II are all quite a bit larger than those obtained in Monte Carlo I.

In Table 8.2 MSE(PTR) is always less than MSE(PT) and MSE(R) for a given α level. Again, the Monte Carlo I estimates of MSE(PTU), MSE(RU), and MSE(PTRU) are not too different from the theoretical values in the first column of Table 8.2. The Monte Carlo II procedure was not carried out in Table 8.2.

In Table 8.3 for $\Delta=2$ and hence $\delta=2.11$, either MSE(R) or MSE(PTR) is lowest for a given α level, with MSE(PT) always being largest. The Monte Carlo I estimates of MSE(PTRU) and MSE(PTU) do not agree as well with MSE(PTR) and MSE(PT) as in the previous two tables. As in Table 8.1, the Monte Carlo II estimates of MSE(PTRU) and MSE(PTU) are much higher than the Monte Carlo I estimates.

The following conclusions can be drawn from Tables 8.1 through 8.3. (1) The estimator PTR is preferable to PT and R for small δ. (2) When

δ increases to a moderate value around 2, either PTR at a high α level (which thus rejects the null hypothesis often) or R is preferable. (3) When δ is large, i.e. greater than 3, the estimators R and PTR are virtually the same and are preferable to PT. (4) In the case of unknown covariance matrix, the estimators PTRU and PTU as calculated in Monte Carlo I seem to have mean square error reasonably close to MSE(PTR) and MSE(PT), respectively.

Table 8.4 (on two pages) provides a comparison of the Bayesian estimators BPM and BN. Using Tables 8.1 through 8.3 with Table 8.4, a comparison of the Bayesian and classical estimators can be made since $(\sigma_y^2,\sigma_x^2) = (6,12)$, $(n,n_x) = (9,18)$, and $\rho = 1/3$ for all four tables.

First, note in Table 8.4 the differences between BPM and BN. For a given population value $\Delta = \mu_y - \mu_x$ and a given $a^2 = V(\tilde{\Delta})$, the bias and variance of BPM remain the same over any specifications of (b_y,α_y^2) since the estimator BPM is not a function of b_y and α_y^2. Hence, for any given row in Table 8.4, the bias and variance of BPM in column 1 can be compared to the bias and variance of BN in the remaining columns under various specifications of (b_y,α_y^2). For example, in rows 1 through 3 of Table 8.4 note that MSE(BPM) > MSE(BN) except for $(b_y,c_y^2) = (2,1)$. From additional numerical results not given in Table 8.4, MSE(BN) > MSE(BPM) for $\Delta = 0$ whenever $|b_y - \mu_y|/\alpha_y > 1.4$. Note that BN has non-zero bias for $b_y \neq 0$, but V(BN) is usually less than V(BPM) and hence MSE(BN) is often less than MSE(BPM).

However, once $|b_y - \mu_y|/\alpha_y$ gets larger than 1.4, then the bias is significant enough to cause MSE(BN) to be larger than MSE(BPM).

Second, in the first four columns of Table 8.4 note that the change from (Δ, a^2) to $(-\Delta, a^2)$ produces the same mean square error for BPM and BN and the same absolute value of the bias. Only the algebraic sign of the bias is affected. Note in columns 1 through 4 that either $b_y = 0$ or b_y is not relevant to the estimator. In columns 5 through 8, however, this is not true. $|B(BN)|$ and MSE(BN) are always larger when Δ is negative. The reason for this is easy to see from an inspection of equation (8.3.5) for B(BN). In Table 8.4, $\mu_y = 0$ and $b_y \geq 0$, so that for $b_y > 0$, the first term in B(BN) is positive. Now if $\Delta > 0$ also, then the contribution to the bias made by $b_y > 0$ and $\Delta > 0$ cancel to some degree in equation (8.3.5). If $\Delta < 0$, however, both terms in equation (8.3.5) are positive, and the bias contributed by each part is additive.

The following general conclusions were obtained from Table 8.4 and some additional computations. One, if $\Delta \geq 0$ and $(b_y - \mu_y) \geq 0$ [or $\Delta \leq 0$ and $(b_y - \mu_y) \leq 0$], then MSE(BPM) > MSE(BN) whenever $|b_y - \mu_y|/\alpha_y > 1.4$ and $\Delta = 0$ or whenever $|b_y - \mu_y|/\alpha_y \geq 2$ and $|\Delta|/a < .7$. In other words, if $|b_y - \mu_y|/\alpha_y$ and $|\Delta|/a$ are approximately equal [recall Δ and $(b_y - \mu_y)$ have the same sign], then the bias effect cancels and BN has the smaller mean square error. It appears there must be a difference of about 1.4 between $|b_y - \mu_y|/\alpha_y$ and $|\Delta|/a$ before BPM has a smaller mean square error. Second, if now $\Delta \leq 0$ and $(b_y - \mu_y) \geq 0$ [or $\Delta \geq 0$ and $(b_y - \mu_y) \leq 0$], the biases do not cancel. In this case MSE(BPM) < MSE(BN) if $|b_y - \mu_y|/\alpha_y > 1.4$, no matter what value $|\Delta|/a$ has. For $|b_y - \mu_y|/\alpha_y \doteq 1$ in the numerical study, MSE(BPM) < MSE(BN) when $|\Delta|/\alpha_y > .50$. For $|b_y - \mu_y|/\alpha_y < 1$, it appears that MSE(BN) < MSE(BPM) no matter what $|\Delta|/a$ is. Hence, the numerical study indicates that for $\Delta \leq 0$ and $(b_y - \mu_y) \geq 0$, MSE(BPM) < MSE(BN) if $|b_y - \mu_y|/\alpha_y \geq 1$ and the sum of $|b_y - \mu_y|/\alpha_y$ and $|\Delta|/a$ is greater than 1.4.

A comparison of the first column of Tables 8.1 through 8.3 with Table 8.4 gives a comparison of the Bayesian estimators BN and BPM to the classical estimators PT, PTR, and R under the assumption of known covariance matrix. For $\Delta = 0$, MSE(R) is greater than all the Bayesian estimators except for two cases under $(b_y, \alpha_y^2) = (2,1)$. Likewise, PT at $\alpha = .50$ and $\alpha = .25$ have higher mean square errors than most of the Bayesian estimators. The estimator PT at $\alpha = .10$ seems equivalent to using BPM. If nothing is known about μ_y, i.e. precise measurement would have to be used on $\tilde{\mu}_y$ in the Bayesian approach, then it seems best to use PTR at a low α-level. If some precise prior information is available on μ_y, then it may be best to use BN since $(b_y, \alpha_y^2) = (0,1)$ and $(0,2)$ yield a value for MSE(BN) which is smaller than or approximately equal to MSE(PTR) for $\alpha = .10$.

For $\Delta = 1$, note that MSE(BPM) is less than the mean square errors for all of the classical estimators. Also, most all values of MSE(BN) are less than the mean square errors of PT, PTR, and R, the exceptions occurring for

$(b_y, \alpha_y^2) = (2,1)$. This, of course, means that $|b_y - \mu_y|/\alpha_y = 2$ and a very poor prior distribution was placed on $\tilde{\mu}_y$. Hence, the Bayesian estimators appear to have lower mean square errors than the classical estimators for $\Delta = 1$ unless a very unfortunate prior is placed upon $\tilde{\mu}_y$.

For $\Delta = 2$, the Bayesian estimators have lower mean square error than PT and PTR at low α-levels, except for a few cases where $(b_y, \alpha_y^2) = (2,1)$. If there is not enough information to warrant a prior distribution on $\tilde{\mu}_y$, then it seems best to use PTR at $\alpha = .50$. However, if a reasonably precise prior can be placed on $\tilde{\mu}_y$, the Bayesian estimators will give smaller mean square errors than the classical estimators.

In this particular numerical study the Bayesian estimators generally appear to have smaller mean square errors than the classical estimators. Two exceptions to this are: (1) when $(b_y - \mu_y) > 0$ and $\Delta < 0$ [or $(b_y - \mu_y) < 0$ and $\Delta > 0$], the biases from the two components are additive and can yield a large bias, (2) if $|b_y - \mu_y|/\alpha_y$ is close to 2 or larger, the large bias causes the mean square error of BN to surpass the mean square error of the classical estimators.

Table 8.1. Bias, variance, and mean square error of PTR, PT, and R for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\rho = 1/3$, $(n, n_x) = (9,18)$, $\Delta = 0$

Estimator		Theoretical	Monte Carlo I	Monte Carlo II
	$\alpha = .50$	0.0000[a]	0.0291	0.0142
		0.5719[b]	0.5808	1.1199
		0.5719[c]	0.5816	1.1200
PTR	$\alpha = .25$	0.0000	0.0155	0.0056
		0.4532	0.5221	1.0753
		0.4532	0.5223	1.0753
	$\alpha = .10$	0.0000	0.0075	-0.0124
		0.2900	0.4623	0.9294
		0.2900	0.4623	0.9296
	$\alpha = .50$	0.0000	0.0311	0.0177
		0.6416	0.5475	1.1780
		0.6416	0.5484	1.1783
PT	$\alpha = .25$	0.0000	0.0231	0.0102
		0.5699	0.4846	1.0499
		0.5699	0.4852	1.0500
	$\alpha = .10$	0.0000	0.0120	-0.0041
		0.4703	0.4297	0.9208
		0.4703	0.4298	0.9208
R		0.0000	0.0044	
		0.6173	0.6584	
		0.6173	0.6584	

[a]Bias

[b]Variance

[c]Mean Square Error

Table 8.2. Bias, variance, and mean square error of PTR, PT, and R for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\rho = 1/3$, $(n,n_x) = (9,18)$, $\Delta = 1$

Estimator		Theoretical	Monte Carlo I
	$\alpha = .50$	-0.0415[a]	-0.0184
		0.5649[b]	0.6374
		0.5666[c]	0.6377
PTR	$\alpha = .25$	-0.1311	-0.0439
		0.5665	0.6273
		0.5838	0.6292
	$\alpha = .10$	-0.2734	-0.0867
		0.5349	0.6137
		0.6096	0.6212
	$\alpha = .50$	-0.0269	-0.0494
		0.6829	0.6342
		0.6836	0.6367
PT	$\alpha = .25$	-0.1129	-0.0856
		0.7146	0.6135
		0.7274	0.6208
	$\alpha = .10$	-0.2565	-0.1244
		0.7077	0.5847
		0.7735	0.6002
R		0.0000	-0.0132
		0.6173	0.6442
		0.6173	0.6444

[a]Bias

[b]Variance

[c]Mean Square Error

Table 8.3. Bias, variance, and mean square error of PTR, PT, and R for $(\sigma_y^2, \sigma_x^2) = (6,12)$, $\rho = 1/3$, $(n, n_x) = (9,18)$, $\Delta = 2$

Estimator		Theoretical	Monte Carlo I	Monte Carlo II
	$\alpha = .50$	-0.0187[a] 0.5853[b] 0.5856[c]	-0.0750 0.7583 0.7640	-0.1095 1.4679 1.4799
PTR	$\alpha = .25$	-0.0753 0.6428 0.6485	-0.2050 0.8703 0.9124	-0.2895 1.6207 1.7045
	$\alpha = .10$	-0.2082 0.7950 0.8384	-0.4561 0.9880 1.1961	-0.6841 1.7311 2.1991
	$\alpha = .50$	-0.0117 0.6913 0.6914	-0.0887 0.7679 0.7758	-0.1162 1.4451 1.4586
PT	$\alpha = .25$	-0.0623 0.7837 0.7876	-0.2357 0.8619 0.9175	-0.3104 1.5893 1.6857
	$\alpha = .10$	-0.1895 0.9610 0.9969	-0.5041 0.9371 1.1191	-0.7150 1.6638 2.1751
R		0.0000 0.6173 0.6173	0.0044 0.6584 0.6584	

[a] Bias

[b] Variance

[c] Mean Square Error

Table 8.4. Bias and mean square error of BN and BPM for $(\sigma_y^2,\sigma_x^2) = (6,12)$, $(n,n_x) = (9,18)$, $\rho = 1/3$

(Δ,a^2)	BPM	BN $(b_y,\alpha_y^2)=(0,1)$	BN $(b_y,\alpha_y^2)=(0,2)$	BN $(b_y,\alpha_y^2)=(0,4)$
(0,1)	0.0000[a] 0.3989[b]	0.0000 0.1832	0.0000 0.2604	0.0000 0.3186
(0,2)	0.0000 0.4606	0.0000 0.1980	0.0000 0.2889	0.0000 0.3599
(0,4)	0.0000 0.5185	0.0000 0.2122	0.0000 0.3157	0.0000 0.3984
(1,1)	-0.2767 0.4755	-0.1875 0.2184	-0.2236 0.3104	-0.2473 0.3798
(-1,1)	0.2767 0.4755	0.1875 0.2184	0.2236 0.3104	0.2473 0.3798
(1,4)	-0.1056 0.5297	-0.0676 0.2169	-0.0824 0.3225	-0.0926 0.4070
(-1,4)	0.1056 0.5297	0.0676 0.2168	0.0824 0.3225	0.0926 0.4070
(2,1)	-0.5534 0.7052	-0.3751 0.3239	-0.4471 0.4603	-0.4946 0.5633
(-2,1)	0.5534 0.7052	0.3751 0.3239	0.4471 0.4603	0.4946 0.5633
(2,4)	-0.2113 0.5632	-0.1352 0.2305	-0.1648 0.3429	-0.1852 0.4327
(-2,4)	0.2113 0.5632	0.1352 0.2305	0.1648 0.3429	0.1852 0.4327

[a] Bias

[b] Mean square error

Table 8.4 (Continued)

(Δ,a^2)	BN $(b_y,\alpha_y^2)=(1,1)$	BN $(b_y,\alpha_y^2)=(1,4)$	BN $(b_y,\alpha_y^2)=(2,1)$	BN $(b_y,\alpha_y^2)=(2.4)$
(0,1)	0.3222 0.2871	0.1062 0.3299	0.6445 0.5986	0.2125 0.3638
(0,2)	0.3443 0.3166	0.1161 0.3734	0.6887 0.6723	0.2321 0.4138
(0,4)	0.3603 0.3420	0.1234 0.4137	0.7205 0.7314	0.2468 0.4594
(1,1)	0.1347 0.2014	-0.1411 0.3385	0.4570 0.3920	-0.0348 0.3198
(-1,1)	0.5098 0.4431	0.3536 0.4436	0.8320 0.8755	0.4598 0.5300
(1,4)	0.2927 0.2979	0.0308 0.3994	0.6530 0.6386	0.1542 0.4222
(-1,4)	0.4278 0.3953	0.2160 0.4451	0.7881 0.8333	0.3394 0.5136
(2,1)	-0.0528 0.1860	-0.3884 0.4695	0.2694 0.2558	-0.2821 0.3982
(-2,1)	0.6973 0.6695	0.6009 0.6797	1.0196 1.2228	0.7071 0.8186
(2,4)	0.2251 0.2629	-0.0618 0.4022	0.5854 0.5549	0.0616 0.4022
(-2,4)	0.4954 0.4576	0.3086 0.4937	0.8557 0.9444	0.4320 0.5851

IX. A BAYESIAN APPROACH TO POOLING PROPORTIONS

A. Specification of Problem

Consider two independent random variables $\tilde{x}_1$ and $\tilde{x}_2$ where $\tilde{x}_1$ follows the binomial distribution with parameters n_1 and p_1, i.e. $b(n_1, p_1)$, and $\tilde{x}_2$ is distributed $b(n_2, p_2)$. It is desired to estimate p_1, with the possibility of pooling $\tilde{x}_1$ and $\tilde{x}_2$ in some manner if p_1 and p_2 are equal. Kale and Bancroft (1967) have discussed this problem from the classical, preliminary test point of view. They approximated the discrete binomial distribution by the normal distribution and then proceeded in a manner similar to the presentation in Chapters III and IV. This chapter discusses the problem from a Bayesian viewpoint, i.e. by assigning a joint prior distribution on $\tilde{p}_1$ and $\tilde{p}_2$ and then finding the posterior distribution of $\tilde{p}_1$. The problem of finding an appropriate joint prior distribution on $(\tilde{p}_1, \tilde{p}_2)$ is discussed in sections B and C, and a new prior distribution possessing several desirable properties is presented in section D. Some numerical work illustrating the use of the prior distribution is presented in section E.

B. Multivariate Beta Distributions

A Beta distribution on $\tilde{p}$ is commonly used when $\tilde{x}$ is distributed $b(n, p)$ and it is desired to estimate p. The Beta distribution, which is the natural conjugate of the binomial distribution, is a two parameter distribution and thus allows selection of the parameters to yield a specified prior mean and variance. Furthermore, the Beta distribution admits a variety of shapes and therefore can represent a wide variety of prior opinions about p.

It is necessary to obtain a joint prior distribution on $(\tilde{p}_1, \tilde{p}_2)$ where $\tilde{p}_1$ and $\tilde{p}_2$ are related so as to allow expression of a prior opinion about the equality of p_1 and p_2. From the previous paragraph, Beta marginals on $\tilde{p}_1$ and $\tilde{p}_2$ seem desirable. Note, however, that independent Beta distributions on $\tilde{p}_1$ and $\tilde{p}_2$ will not suffice because the posterior mean of $\tilde{p}_1$ will then be independent of $\tilde{x}_2$. Hence, a bivariate distribution defined on the unit square with Beta marginals seems to be a likely candidate.

A review of the literature on multivariate Beta distributions reveals rather rapidly that this term is not used uniformly by all authors. Wilks (1962) and Mosimann (1962) say that the Dirichlet distribution is the multivariate analogue of the Beta distribution. The density of the Dirichlet distribution in the k-variate case is

$$f(x_1,\ldots,x_k) = \frac{\Gamma(\upsilon_1+\ldots+\upsilon_{k+1})}{\Gamma(\upsilon_1)\ldots\Gamma(\upsilon_{k+1})}\, x_1^{\upsilon_1-1}\ldots x_k^{\upsilon_k-1}(1-x_1-\ldots-x_k)^{\upsilon_{k+1}-1}$$

$$\text{for } x_i \geq 0 \text{ and } \sum_{i=1}^{k} x_i \leq 1 \text{ and } \upsilon_i > 0$$

$$f(x_1,\ldots x_k) = 0 \text{ elsewhere.} \tag{9.2.1}$$

The Dirichlet distribution is the natural conjugate of the multinomial distribution, and all the marginals of the Dirichlet distribution are Beta. However, if a Dirichlet prior were taken on $\tilde{p}_1$ and $\tilde{p}_2$ in the pooling problem, the sum of the random variables $\tilde{p}_1$ and $\tilde{p}_2$ would have to be between zero and one in this bivariate case. This restriction cannot be made since it may happen, for example, that $p_1 = p_2 = 3/4$. Thus, the Dirichlet distribution is not a suitable prior distribution here.

Olkin and Rubin (1964, 1966) discuss several types of multivariate Beta distributions, including the Dirichlet and inverted Dirichlet distributions. In equations (3.6) and (3.7) of Olkin and Rubin (1964, p. 263), a bivariate distribution on $\tilde{u}_1$ and $\tilde{u}_2$ is defined to be positive over Quadrant I. If now the transformation $\tilde{p}_i = \tilde{u}_i/(1+\tilde{u}_i)$, i = 1, 2, is applied to their density, the resultant bivariate density on $\tilde{p}_1$ and $\tilde{p}_2$ is defined on the unit square and can be shown to have Beta marginals, where $\tilde{p}_1$ and $\tilde{p}_2$ are dependent. Furthermore, the same bivariate density can also be obtained by considering a slight generalization of the bivariate F distribution on $\tilde{F}_1$ and $\tilde{F}_2$ as discussed by Krishnaiah (1965) and Krishnaiah and Armitage (1964) and then making the transformation $p_i = \left[n_i F_i/m\right] \Big/ \left[1+n_i F_i/m\right]$, where F_i has (n_i, m) df for i = 1, 2. Although this resultant bivariate distribtuion on $\tilde{p}_1$ and $\tilde{p}_2$ has some desirable properties, the assumption that $E(\tilde{p}_1) = E(\tilde{p}_2)$, paralleling the discussion of $E(\tilde{\mu}_x) = E(\tilde{\mu}_y)$ in Chapter VII, implies that the marginals on $\tilde{p}_1$ and $\tilde{p}_2$ are identical. This is not desirable, since it is most likely that $V(\tilde{p}_2) > V(\tilde{p}_1)$. In addition, this bivariate density does not combine nicely with the binomial density. [In equation (3.5) of Olkin and Rubin (1964) the constant C should be

$$\frac{\pi^{-\frac{kp(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left[\frac{n-i+1}{2}\right]}{\prod_{\alpha=0}^{k} \prod_{i=1}^{p} \Gamma\left[\frac{n_\alpha-i+1}{2}\right]}$$

rather than

$$\pi^{-\frac{kp(p-1)}{4}} \prod_{i=1}^{p} \prod_{\alpha=0}^{k} \frac{\Gamma\left[\frac{n_\alpha-i+1}{2}\right]}{\Gamma\left[\frac{n-i+1}{2}\right]}$$

as stated in the paper.]

Khatri (1965), Geisser (1965b) and Kshirsagar (1961) discuss multivariate Beta distributions in which the density is for a random matrix L and is of the form $k|L|^{c_1} |I-L|^{c_2}$, where I is the identity matrix. Mauldon (1959) investigates several generalizations of the Beta distribution and calls the Dirichlet distribution the "basic" Beta distribution. Foster and Rees (1957) present another kind of generalized Beta distribution where the restriction on p_1 and p_2 is $0 < p_1 \leq p_2 < 1$. This density is also not adaptable to the problem at hand.

Most all of the distributions which are called multivariate Beta distributions are "derived" distributions, i.e. they are distributions of functions of random variables. In most cases, the original random variables are normally distributed. What is needed in the proportion pooling problem is a distribution motivated by multivariate data, which, unfortunately, seems to be non-existent.

In conclusion, then, none of the existing multivariate or bivariate Beta distributions are appropriate to the problem of pooling proportions from a Bayesian viewpoint.

C. Approaches to Finding a Prior Distribution

Mentioned in this section are several intuitive suggestions for a prior on $\tilde{p}_1$ and $\tilde{p}_2$, followed by reasons why they are unsatisfactory.

Define now $D = p_1 - p_2$. Clearly, it is impossible to have the distributions of $\tilde{p}_1$ and $\tilde{D}$ be independent as was done in Chapter VII because once $\tilde{p}_1 = p_1$, then $\tilde{D}$ can no longer be in the interval $(-1, 1)$, but must lie in the restricted interval (p_1-1, p_1). Likewise, if $\tilde{D} = D$, then $\tilde{p}_1$ must lie in the intersection of the intervals $(D, D+1)$ and $(0, 1)$.

Let $\tilde{p}_1$ follow the Beta distribution with parameters α_1 and β_1, i.e. $\beta(\alpha_1, \beta_1)$. If now the distribution of $\tilde{D} \mid p_1$ is uniform over (p_1-1, p_1), then it can be shown that the resultant joint prior on $\tilde{p}_1$ and $\tilde{D}$ is equivalent to a joint prior distribution on $\tilde{p}_1$ and $\tilde{p}_2$ where $\tilde{p}_1$ is distributed $\beta(\alpha_1, \beta_1)$, $\tilde{p}_2$ is uniform over $(0, 1)$, and $\tilde{p}_1$ and $\tilde{p}_2$ are independent. The independence of $\tilde{p}_1$ and $\tilde{p}_2$ makes this joint prior distribution unsatisfactory. For the same reasons, taking $\tilde{p}_1$ to be distributed $\beta(\alpha_1, \beta_1)$ and using precise measurement on $\tilde{p}_2 \mid p_1$ or $\tilde{D} \mid p_1$ is not satisfactory.

Let now $\tilde{p}_1$ be distributed $\beta(\alpha_1, \beta_1)$, and let $\tilde{D} \mid p_1$ have the Beta distribution over $(p_1-1,\ p_1)$ with parameters α_2 and β_2, i.e. a type I distribution (Cramèr, 1946, p. 249). As in the previous paragraph, it can be shown that this is equivalent to $\tilde{p}_1$ and $\tilde{p}_2$ being independently distributed and hence is not a useful prior.

As an alternative approach define $\Delta = p_2 - p_1$, so that $0 < p_1 + \Delta < 1$. It now seems plausible to consider a Dirichlet distribution on $\tilde{p}_1$ and $\tilde{\Delta}$, except that the Dirichlet distribution is positive only over the triangular region R_1 in Figure 9.1, and the joint distribution of $\tilde{\Delta}$ and $\tilde{p}_1$ must be positive over the region $R_1 \cup R_2$ (for this chapter).

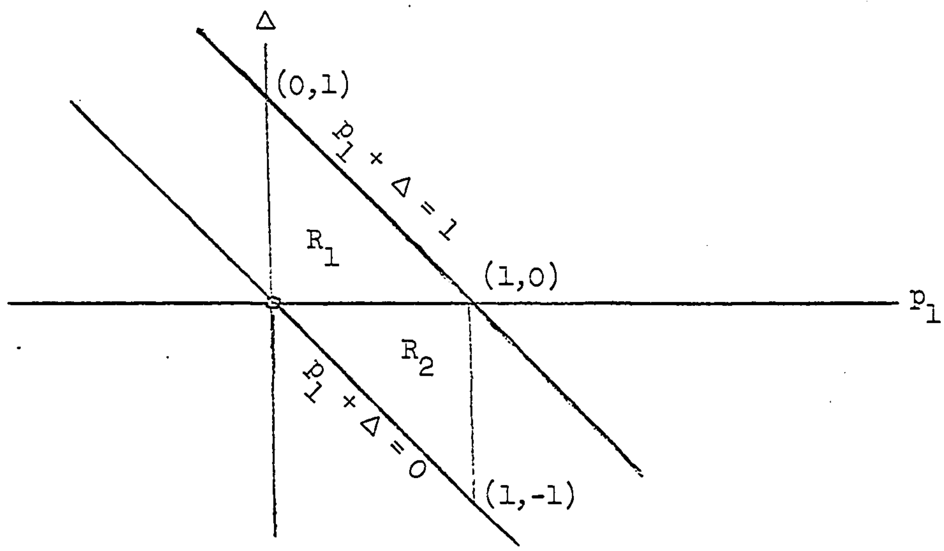

Figure 9.1. Region $R_1 \cup R_2$ where $g(p_1, \Delta) > 0$

Since it is desired to have $E(\tilde{\Delta}) = 0$, then it is possible to place one half of the regular Dirichlet density over the region R_1, and place the remaining half of the probability over R_2 in such a manner that

$\Pr\left[a < \tilde{\Delta} < b\right] = \Pr\left[-b < \tilde{\Delta} - a\right]$ for $a > 0$ and $b > 0$. However, this procedure implies $E(\tilde{p}_1) = 1/2$, which is too restrictive. Nevertheless, similar reasoning and some juggling lead to the bivariate prior on $(\tilde{p}_1, \tilde{\Delta})$ presented in section D.

D. A Bivariate Prior on $\tilde{p}_1$ and $\tilde{\Delta}$

1. The bivariate prior and its properties

Define now the prior density function $g(p_1, \Delta)$ as

$$g(p_1, \Delta) = g_1(p_1, \Delta) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)\, p_1^{\alpha_1-1}\, \Delta^{\alpha_2-1}\, (1-p_1-\Delta)^{\alpha_3-1}}{2\Gamma(\alpha_1)\,\Gamma(\alpha_2)\,\Gamma(\alpha_3)} \text{ over } R_1$$

$$g(p_1,\Delta) = g_2(p_1,\Delta) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)(1-p_1)^{\alpha_2+\alpha_3-1}\,(-\Delta)^{\alpha_2-1}\,(p_1+\Delta)^{\alpha_1-\alpha_2-1}}{2\Gamma(\alpha_2+\alpha_3)\,\Gamma(\alpha_2)\,\Gamma(\alpha_1-\alpha_2)}$$

over R_2

$$g(p_1,\Delta) = 0 \text{ elsewhere} \tag{9.4.1}$$

where the regions R_1 and R_2, illustrated in Figure 9.1, are

$$\begin{aligned} R_1 &= \left[(p_1,\Delta)\colon 0 < p_1 + \Delta < 1,\ \Delta \geq 0,\ p_1 \geq 0\right] \\ R_2 &= \left[(p_1,\Delta)\colon 0 < p_1 + \Delta < 1,\ \Delta < 0,\ p_1 > 0\right] \end{aligned} \tag{9.4.2}$$

and the parameters α_1, α_2, α_3 satisfy

$$\alpha_1 > \alpha_2 > 0, \quad \alpha_3 > 0. \tag{9.4.3}$$

This density can be easily shown to integrate out to one by integrating first with respect to Δ and applying the transformation $z = \Delta/(1-p_1)$ in R_1 and $y = -\Delta/p_1$ in R_2.

The marginal prior distribution on $\tilde{\Delta}$, obtained by integrating $g(p_1,\Delta)$ with respect to p_1, is symmetric about zero with density

$$h(\Delta) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)\ |\Delta|^{\alpha_2-1}\ (1-|\Delta|)^{\alpha_1+\alpha_3-1}}{2\Gamma(\alpha_1+\alpha_3)\ \Gamma(\alpha_2)} \text{ for } -1 < \Delta < 1$$

$$h(\Delta) = 0 \text{ elsewhere.} \tag{9.4.4}$$

The prior mean and variance of $\tilde{\Delta}$ are obtained as

$$E(\tilde{\Delta}) = 0$$

$$V(\tilde{\Delta}) = \frac{\alpha_2(\alpha_2+1)}{(\alpha_1+\alpha_2+\alpha_3)(\alpha_1+\alpha_2+\alpha_3+1)}\ . \tag{9.4.5}$$

The marginal prior distribution on $\tilde{p}_1$ is $\beta(\alpha_1,\ \alpha_2 + \alpha_3)$, i.e.

$$k(p_1) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_2+\alpha_3)\Gamma(\alpha_1)}\, p_1^{\alpha_1-1}\ (1-p_1)^{\alpha_2+\alpha_3-1}, 0 < p_1 < 1$$

$$k(p_1) = 0 \text{ elsewhere.} \tag{9.4.6}$$

Also,

$$E(\tilde{p}_1) = \alpha_1/(\alpha_1+\alpha_2+\alpha_3)$$

$$V(\tilde{p}_1) = \frac{\alpha_1(\alpha_2+\alpha_3)}{(\alpha_1+\alpha_2+\alpha_3)^2(\alpha_1+\alpha_2+\alpha_3+1)}\ . \tag{9.4.7}$$

It is also easily shown that

$$\mathrm{cov}(\tilde{p}_1, \tilde{\Delta}) = \frac{-\alpha_2}{2(\alpha_1+\alpha_2+\alpha_3+1)(\alpha_1+\alpha_2+\alpha_3)} \tag{9.4.8}$$

The covariance of $\tilde{p}_1$ and $\tilde{p}_2$ is then obtained as

$$\mathrm{cov}\ (\tilde{p}_1, \tilde{p}_2) = V(\tilde{p}_1) + \mathrm{cov}(\tilde{p}_1, \tilde{\Delta}). \tag{9.4.9}$$

The conditional distribution of $\tilde{p}_1$, given $\tilde{\Delta} = \Delta$, is obtained as

$$g(p_1 \mid \Delta) = g_1(p_1 \mid \Delta) = \frac{\Gamma(\alpha_1+\alpha_3)\ p_1^{\alpha_1-1}(1-p_1-\Delta)^{\alpha_3-1}}{\Gamma(\alpha_1)\ \Gamma(\alpha_3)(1-\Delta)^{\alpha_1+\alpha_3-1}}$$

$$\text{for } \Delta > 0,\ 0 < p_1 < 1 - \Delta$$

$$g(p_1 \mid \Delta) = g_2(p_1 \mid \Delta) = \frac{\Gamma(\alpha_1+\alpha_3)(1-p_1)^{\alpha_2+\alpha_3-1}(p_1+\Delta)^{\alpha_1-\alpha_2-1}}{\Gamma(\alpha_2+\alpha_3)\Gamma(\alpha_1-\alpha_2)(1+\Delta)^{\alpha_1+\alpha_3-1}}$$

$$\text{for } \Delta < 0,\ -\Delta < p_1 < 1$$

$$g(p_1 \mid \Delta) = 0 \text{ elsewhere.} \tag{9.4.10}$$

Thus, $g_1(p_1 \mid \Delta)$ and $g_2(p_1 \mid \Delta)$ are both Type I distributions, i.e. "generalized" Beta distributions. Using the marginal of $\tilde{p}_1$ in equation (9.4.6), the conditional distribution of $\tilde{\Delta} \mid p_1$ is found to be

$$h(\Delta|p_1) = h_1(\Delta|p_1) = \frac{\Gamma(\alpha_2+\alpha_3)\ \Delta^{\alpha_2-1}(1-p_1-\Delta)^{\alpha_3-1}}{2\Gamma(\alpha_2)\ \Gamma(\alpha_3)(1-p_1)^{\alpha_2+\alpha_3-1}},\quad 0 \leq \Delta < 1-p,$$

$$h(\Delta \mid p_1) = h_2(\Delta \mid p_1) = \frac{\Gamma(\alpha_1)(-\Delta)^{\alpha_2-1}(p_1+\Delta)^{\alpha_1-\alpha_2-1}}{2\Gamma(\alpha_2)\,\Gamma(\alpha_1-\alpha_2)\,p_1^{\alpha_1-1}}, -p_1 < \Delta < 0$$

$$h(\Delta \mid p_1) = 0 \text{ elsewhere.} \tag{9.4.11}$$

Note that the functions $h_1(\Delta \mid p_1)$ and $h_2(\Delta \mid p_1)$ are each one-half of a generalized Beta distribution.

Hence, the prior bivariate density presented in equation (9.4.1) has the following desirable properties: 1) the marginal of $\tilde{p}_1$ is a Beta distribution; 2) the marginal of $\tilde{\Delta}$ is symmetric about zero; 3) there are three parameters to vary, so that, for example, the prior mean and variance of $\tilde{p}_1$ and the prior variance of $\tilde{\Delta}$ can be specified and then α_1, α_2, and α_3 can be chosen accordingly; 4) the covariance of $\tilde{p}_1$ and $\tilde{\Delta}$ can be made small by selecting α_2 small.

2. <u>The posterior distribution of $\tilde{p}_1$</u>

Consider now how the prior density $g(p_1, \Delta)$ in equation (9.4.1) combines with the conditional data distribution $f(x_1, x_2 \mid p_1, \Delta)$ where

$$f(x_1,x_2 \mid p_1,\Delta) = \binom{n_1}{x_1}\binom{n_2}{x_2} p_1^{x_1}(1-p_1)^{n_1-x_1}(\Delta+p_1)^{x_2}(1-\Delta-p_1)^{n_2-x_2} \tag{9.4.12}$$

Then the joint distribution $q(x_1,x_2,p_1,\Delta)$ of $\tilde{x}_1,\tilde{x}_2,\tilde{p}_1,\tilde{\Delta}$ is given by

$$q(x_1,x_2,p_1,\Delta) = q_1(x_1,x_2,p_1,\Delta) = f(x_1,x_2|p_1,\Delta)\ g_1(p_1,\Delta)$$

$$\text{for } (p_1,\Delta)\epsilon R_1,\ x_1 = 0, 1, \ldots, n_1,\ x_2 = 0, 1,\ldots,n_2.$$

$$q(x_1,x_2,p_1,\Delta) = q_2(x_1,x_2,p_1,\Delta) = f(x_1,x_2|p_1,\Delta)\ g_2(p_1,\Delta)$$

$$\text{for } (p_1,\Delta)\epsilon R_2,\ x_1 = 0, 1, \ldots, n_1,\ x_2=0, 1,\ldots,n_2.$$

$$q(x_1,x_2,p_1,\Delta) = 0 \text{ elsewhere.} \tag{9.4.13}$$

Using the binomial theorem to expand terms like $(\Delta+p_1)^{x_2}$ and $[1-(\Delta+p_1)]^{n_2-x_2}$, the joint distribution $t(x_1,x_2,p_1)$ can be obtained as

$$t(x_1,x_2,p_1) = K_1 \sum_{k=0}^{x_2} Y_1(k) + K_2 \sum_{k=0}^{n_2-x_2} Y_2(k) \tag{9.4.14}$$

$$\text{for } 0 < p_1 < 1,\ x_1 = 0, 1, \ldots, n_1,\ x_2 = 0, 1,\ldots,n_2.$$

$$t(x_1,x_2,p_1) = 0 \text{ elsewhere,}$$

where

$$K_1 = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{2\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \binom{n_1}{x_1}\binom{n_2}{x_2} \tag{9.4.15}$$

$$K_2 = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{2\Gamma(\alpha_2+\alpha_3)\Gamma(\alpha_2)\Gamma(\alpha_1-\alpha_2)} \binom{n_1}{x_1}\binom{n_2}{x_2} \tag{9.4.16}$$

$$Y_1(k) = \binom{x_2}{k} p_1^{x_1+x_2+\alpha_1-k-1} (1-p_1)^{n_1-x_1+n_2-x_2+\alpha_2+\alpha_3+k-1}$$

$$\times \frac{\Gamma(\alpha_2+k)\Gamma(n_2-x_2+\alpha_3)}{\Gamma(n_2-x_2+\alpha_2+\alpha_3+k)} \tag{9.4.17}$$

$$Y_2(k) = \binom{n_2-x_2}{k} p_1^{x_1+x_2+\alpha_1+k-1} (1-p_1)^{n_1-x_1+n_2-x_2+\alpha_2+\alpha_3-k-1}$$

$$\times \frac{\Gamma(\alpha_2+k)\Gamma(x_2+\alpha_1-\alpha_2)}{\Gamma(x_2+\alpha_1+k)} \tag{9.4.18}$$

The unconditional distribution of $\tilde{x}_1$ and $\tilde{x}_2$, i.e. $w(x_1, x_2)$, is obtained by integrating equation (9.4.14) with respect to p_1, which yields

$$w(x_1, x_2) = K_1 \sum_{k=0}^{x_2} Z_1(k) + K_2 \sum_{k=0}^{n_2-x_2} Z_2(k), \tag{9.4.19}$$

$$\text{for } x_1 = 0, 1, \ldots, n_1 \text{ and } x_2 = 0, 1, \ldots, n_2$$

$$w(x_1, x_2) = 0 \text{ elsewhere,}$$

where

$$Z_1(k) = \binom{x_2}{k} \frac{\Gamma(x_1+x_2+\alpha_1-k)\Gamma(n_1-x_1+n_2-x_2+\alpha_2+\alpha_3+k)\Gamma(\alpha_2+k)\Gamma(n_2-x_2+\alpha_3)}{\Gamma(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)\Gamma(n_2-x_2+\alpha_2+\alpha_3+k)} \tag{9.4.20}$$

and

$$Z_2(k) = \binom{n_2-x_2}{k} \frac{\Gamma(x_1+x_2+\alpha_1+k)\Gamma(n_1-x_1+n_2-x_2+\alpha_2+\alpha_3-k)\Gamma(\alpha_2+k)\Gamma(x_2+\alpha_1-\alpha_2)}{\Gamma(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)\Gamma(x_2+\alpha_1+k)} \tag{9.4.21}$$

Thus, the posterior distribution of $\tilde{p}_1$ is obtained as

$$d(p_1|x_1,x_2) = t(x_1,x_2,p_1)/ \left[w(x_1,x_2)\right]$$

$$\text{for } 0 < p_1 < 1,\ x_1 = 0, 1, \ldots, n_1,\ x_2 = 0, 1, \ldots, n_2$$

$$d(p_1|x_1,x_2) = 0 \text{ elsewhere.} \tag{9.4.22}$$

Note that the denominator of $d(p_1|x_1,x_2)$ is a sum of multiples of gamma functions whose arguments involve the sample results and the known parameters of the data and prior distributions. The numerator of $d(p_1|x_1,x_2)$ is a sum of multiples of gamma functions and powers of p_1 and $(1-p_1)$.

A mathematical expression for the posterior mean is obtained very easily as

$$E(\tilde{p}_1|x_1,x_2) = \frac{K_1 \sum\limits_{k=0}^{x_2} \frac{(x_1+x_2+\alpha_1-k)Z_1(k)}{(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)} + K_2 \sum\limits_{k=0}^{n_2-x_2} \frac{(x_1+x_2+\alpha_1+k)Z_2(k)}{(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)}}{K_1 \sum\limits_{k=0}^{x_2} Z_1(k) + K_2 \sum\limits_{k=0}^{n_2-x_2} Z_2(k)} \tag{9.4.23}$$

where K_1, K_2, $Z_1(k)$, and $Z_2(k)$ are defined in equations (9.4.15), (9.4.16), (9.4.20), and (9.4.21), respectively. The posterior expectation of $\tilde{p}_1^2$ can likewise be easily obtained as

$$E(\tilde{p}_1^2|x_1,x_2) = \frac{K_1 \sum\limits_{k=0}^{x_2} a_1(k)\ Z_1(k) + K_2 \sum\limits_{k=0}^{n_2-x_2} a_2(k)\ Z_2(k)}{K_1 \sum\limits_{k=0}^{x_2} Z_1(k) + K_2 \sum\limits_{k=0}^{n_2-x_2} Z_2(k)} \tag{9.4.24}$$

where

$$a_1(k) = \frac{(x_1+x_2+\alpha_1-k+1)(x_1+x_2+\alpha_1-k)}{(n_1+n_2+\alpha_1+\alpha_2+\alpha_3+1)(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)} \tag{9.4.25}$$

and

$$a_2(k) = \frac{(x_1+x_2+\alpha_1+k+1)(x_1+x_2+\alpha_1+k)}{(n_1+n_2+\alpha_1+\alpha_2+\alpha_3+1)(n_1+n_2+\alpha_1+\alpha_2+\alpha_3)} \ . \tag{9.4.26}$$

From equations (9.4.23) and (9.4.24), the posterior variance of $\tilde{p}_1$ can be obtained.

After some algebra, it can be shown that for $n_2 = x_2 = 0$, which corresponds to no sample from $b(n_2, p_2)$, the posterior mean and variance of $\tilde{p}_1$ from equations (9.4.23) and (9.4.24) are

$$\begin{aligned} E(\tilde{p}_1 | x_1, x_2=n_2=0) &= (x_1+\alpha_1)/(n_1+\alpha_1+\alpha_2+\alpha_3) \\ V(\tilde{p}_1 | x_1, x_2=n_2=0) &= \frac{(x_1+\alpha_1)(n_1-x_1+\alpha_2+\alpha_3)}{(n_1+\alpha_1+\alpha_2+\alpha_3)^2(n_1+\alpha_1+\alpha_2+\alpha_3+1)} \end{aligned} \tag{9.4.27}$$

This same result is obtained if a Beta prior with parameters α_1 and $(\alpha_2+\alpha_3)$ is placed on $\tilde{p}_1$ and $\tilde{x}_1$ is distributed $b(n_1, p_1)$. Recall that the marginal prior on $\tilde{p}_1$ is $\beta(\alpha_1, \alpha_2+\alpha_3)$. Thus, for $n_2 = x_2 = 0$, the posterior mean and variance of $\tilde{p}_1$ reduce to the case where observations are available on $\tilde{x}_1$ only and the usual natural conjugate Bayesian analysis is carried out.

E. Some Numerical Results

Although the prior $g(p_1, \Delta)$ presented in equation (9.4.1) has some desirable properties, it is not obvious from inspection of equations (9.4.23) and (9.4.24) how the posterior mean and variance of $\tilde{p}_1$ are affected by the prior distribution and the sample data. In an effort to discover some of the properties of the joint prior $g(p_1, \Delta)$ and the posterior distribution

$d(p_1|x_1,x_2)$, a numerical study was done.

In the numerical study, different joint priors were produced by considering all eight combinations of $\alpha_1 = 1.25, 2.00$; $\alpha_2 = .25, .75$; and $\alpha_3 = 1.00, 2.00$. All of these combinations produced priors on $\tilde{p}_1$ which were rather diffuse or flat, with slight skewness to either the right or left. Also considered were $(\alpha_1, \alpha_2, \alpha_3) = (.5, .25, 4.00)$ and $(\alpha_1, \alpha_2, \alpha_3) = (6.00, .50, .50)$ which produced priors on $\tilde{p}_1$ extremely skewed to the right and left, respectively. The various sample outcomes investigated were $(n_1, n_2, x_1, x_2) = (20, 20, 10, 10)$, $(20, 20, 15, 2)$, $(20, 20, 10, 5)$, $(10, 10, 5, 5)$, $(10, 10, 7, 1)$, and $(10, 10, 5, 2)$. The sample result $(20, 20, 10, 10)$ supports the prior opinion that $p_1 = p_2$; the sample result $(20, 20, 15, 2)$ does not yield support to the prior opinion that p_1 and p_2 are equal; and the sample result $(20, 20, 10, 5)$ offers no decisive support or rejection to the equality of p_1 and p_2. The remaining three sample results have the same properties as the first three, but are only half as large. This offers a comparison of the effect of sample size upon the posterior distribution of $\tilde{p}_1$.

For each specification of $(\alpha_1, \alpha_2, \alpha_3)$ and (n_1, n_2, x_1, x_2), the parameters of the prior distribution $g(p_1, \Delta)$ were calculated, i.e. $E(\tilde{p}_1)$, $V(\tilde{p}_1)$, $V(\tilde{\Delta})$, $\text{cov}(\tilde{p}_1, \tilde{\Delta})$, $\text{cov}(\tilde{p}_1, \tilde{p}_2)$, etc., and the marginal prior distribution of $\tilde{p}_1$ was graphed. The parameters $E(\tilde{p}_1 \mid x_1, x_2)$ and $V(\tilde{p}_1 \mid x_1, x_2)$ of the posterior distribution were calculated, and the posterior distribution of $\tilde{p}_1$ was graphed. A comparison of the graph of the prior marginal $k(p_1)$ to the posterior distribution $d(p_1 \mid x_1, x_2)$ indicates how the sample results change the prior distribution on $\tilde{p}_1$.

Some of the results which were obtained are given in Table 9.2 and Figures 9.1, 9.2, and 9.3. In Figure 9.1 note that the prior marginal on $\tilde{p}_1$ is rather diffuse with prior mean $E(\tilde{p}_1) = .5$ and prior variance $V(\tilde{p}_1) = 0.07143$. (These figures are obtained from Table 9.2.) The sample result $(n_1, n_2, x_1, x_2) = (20, 20, 10, 10)$ strongly supports the prior opinion that $p_1 = p_2 = .5$, and hence from Table 9.1 the posterior mean and variance of $\tilde{p}_1$ are $E(\tilde{p}_1 \mid x_1 = 10, x_2 = 10) = 0.5$ and $V(\tilde{p}_1 \mid x_1 = 10, x_2 = 10) = 0.00689$. Note that the posterior variance is one-tenth of the prior variance. The sample result $(n_1, n_2, x_1, x_2) = (20, 20, 15, 2)$ does not support the prior opinion that $p_1 = p_2 = .5$. In this case the posterior mean is $E(\tilde{p}_1 \mid x_1 = 15, x_2 = 2) = .7057$, which is between the prior mean $E(\tilde{p}_1) = .50$ and the sample mean $x_1/n_1 = .75$. Note that the posterior variance $V(\tilde{p}_1 \mid x_1 = 15, x_2 = 2) = 0.00948$ is lower than the prior variance of $\tilde{p}_1$, but not as low as when $(n_1, n_2, x_1, x_2) = (20, 20, 10, 10)$. The sample result $(n_1, n_2, x_1, x_2) = (20, 20, 10, 5)$ does not clearly confirm or discredit the prior opinion that $p_1 = p_2 = .5$, and the posterior mean and variance for this case are obtained as $E(\tilde{p}_1 \mid x_1 = 10, x_2 = 5) = .4319$ and $V(\tilde{p}_1 \mid x_1 = 10, x_2 = 5) = 0.00988$. Note that the sample mean $x_1/n_1 = .5$ and the prior mean are equal, but the posterior mean is less than .5 due to the influence of x_2. All three posterior distributions in Figure 9.1 are fairly symmetrical and could possibly be approximated by normal or Beta distributions.

Consider now the same prior distribution $g(p_1, \Delta)$ as used in Figure 9.1, but change (n_1, n_2, x_1, x_2) from (20, 20, 10, 10), (20, 20, 15, 2), and (20, 20, 10, 5) to (10, 10, 5, 5), (10, 10, 7, 1), and (10, 10, 5, 2). The difference in the posterior means and variances are given in Table 9.1 below. For the sample result (n_1, n_2, x_1, x_2) = (10, 10, 5, 5) the

Table 9.1. Posterior mean and variance for $(\alpha_1,\alpha_2,\alpha_3) = (1.25,.25,1.00)$

(n_1,n_2,x_1,x_2)	$E(\tilde{p}_1 \mid x_1,x_2)$	$V(\tilde{p}_1 \mid x_1,x_2)$
(20, 20, 10, 10)	0.50000	0.00689
(10, 10, 5, 5)	0.50000	0.01246
(20, 20, 15, 2)	0.70570	0.00948
(10, 10, 7, 1)	0.59746	0.02226
(20, 20, 10, 5)	0.43187	0.00988
(10, 10, 5, 2)	0.41326	0.01568

posterior mean remained the same, but the posterior variance doubled due to having a sample size half as large. For (n_1, n_2, x_1, x_2) = (10, 10, 7, 1), the posterior mean is almost .60, which is between the prior mean of .50 and the sample mean of .70. Again, the posterior variance is more than twice as large as the posterior variance of the larger sample. For the sample result (n_1, n_2, x_1, x_2) = (10, 10, 5, 2), the posterior mean is

0.41326, compared to 0.43187 for the larger sample. Perhaps x_2 is exerting a larger effect when the sample is smaller. The posterior variance is about $1\frac{1}{2}$ times larger for the smaller sample. The graphs of the posterior distributions for the smaller samples (not illustrated here) are very similar to the graphs in Figure 9.1, with the exception that the larger posterior variances make the graphs less concentrated around the posterior mean.

For all other diffuse priors which were considered, similar results to those presented in Figure 9.1 were obtained.

Figure 9.2 represents a case where the prior information about p_1 is not diffuse. The prior mean is $E(\tilde{p}_1) = 0.105$ and the prior variance is $V(\tilde{p}_1) = 0.01638$. Note that the prior variance of $\tilde{p}_1$ is much less in Figure 9.2 than in Figure 9.1. None of the three sample results $(n_1, n_2, x_1, x_2) =$ (20, 20, 10, 10), (20, 20, 15, 2), (20, 20, 10, 5) strongly support the prior opinion that $p_1 = p_2 = 0.105$. The posterior distributions have the same general shape as in Figure 9.1. The posterior means in Figure 9.2 are all lower than the sample means x_1/n_1, but they are more dominated by the sample data than by the prior mean of $\tilde{p}_1$. Hence, it appears that if the sample data are in opposition to the prior opinions about p_1 and p_2, then the posterior distribution is dominated by the sample data. Of course, the sample dominance probably decreases as the sample sizes n_1 and n_2 decrease.

Figure 9.3 gives another example of a non-diffuse prior on $\tilde{p}_1$ with prior mean $E(\tilde{p}_1) = 0.857$ and prior variance $V(\tilde{p}_1) = 0.01531$. Again, none of the sample results strongly support the prior opinion that $p_1 = p_2 = 0.857$, and the posterior distribution is dominated by the sample data.

F. Recommendations on the Use of This Prior

The numerical study reported in section E points to the following conclusions. (1) The posterior distribution of $\tilde{p}_1$ is fairly symmetrical around the posterior mean and could probably be approximated by a normal or Beta distribution. (2) In cases where the prior mean $E(\tilde{p}_1)$ agrees closely with sample results x_1/n_1 and x_2/n_2, the posterior mean is in close agreement with all three quantities. (3) In cases where the prior information is widely divergent from the sample results, the posterior distribution of $\tilde{p}_1$ seems to be heavily influenced by the sample.

There are several other topics, however, which need further investigation. One of these is a further study of the influence of x_2 upon the posterior mean when x_1/n_1 and x_2/n_2 are widely divergent. Ideally, x_2 should contribute little to the posterior mean in such a case since the sample indicates $p_1 \neq p_2$. Also, it would be desirable to find an approximation for the posterior mean and variance, since the formulas in equations (9.4.23) and (9.4.24) are rather cumbersome. Another topic worth investigating is how the prior correlation of $\tilde{p}_1$ and $\tilde{p}_2$ affects the relative weight given to x_1 and x_2 in the calculation of the posterior mean.

Until these further topics have been investigated, it is difficult to give recommendations on the choice of α_1, α_2, α_3 so as to express certain prior opinions other than the obvious method of specifying three of the parameters of the joint prior distribution [such as $E(\tilde{p}_1)$, $V(\tilde{p}_1)$, $V(\tilde{\Delta})$] and then solving for the values of α_1, α_2, α_3 which yield this specification.

Table 9.2. A tabular description of Figures 9.1, 9.2, and 9.3

Prior and posterior properties	Figure 9.1	Figure 9.2	Figure 9.3
$(\alpha_1,\alpha_2,\alpha_3)$	(1.25,.25,1.00)	(.50,.25,4.00)	(6.00,.50,.50)
$E(\tilde{p}_1) = E(\tilde{p}_2)$	0.50000	0.10500	0.85700
$V(\tilde{p}_1)$	0.07143	0.01638	0.01531
$V(\tilde{p}_2)$	0.07857	0.01867	0.01977
$V(\tilde{\Delta})$	0.03571	0.01144	0.01339
$\mathrm{corr}(\tilde{p}_1,\tilde{p}_2)$	0.76277	0.67498	0.62325
$\mathrm{corr}(\tilde{p}_1,\tilde{\Delta})$	0.28284	0.33431	0.31180
$E(\tilde{p}_1 \mid x_1 = 10, x_2 = 10)$	0.50000	0.45107	0.56603
$V(\tilde{p}_1 \mid x_1 = 10, x_2 = 10)$	0.00689	0.00621	0.00620
$E(\tilde{p}_1 \mid x_1 = 15, x_2 = 2)$	0.70570	0.61824	0.71425
$V(\tilde{p}_1 \mid x_1 = 15, x_2 = 2)$	0.00948	0.00959	0.00951
$E(\tilde{p}_1 \mid x_1 = 10, x_2 = 5)$	0.43187	0.37019	0.49624
$V(\tilde{p}_1 \mid x_1 = 10, x_2 = 5)$	0.00988	0.00745	0.00812

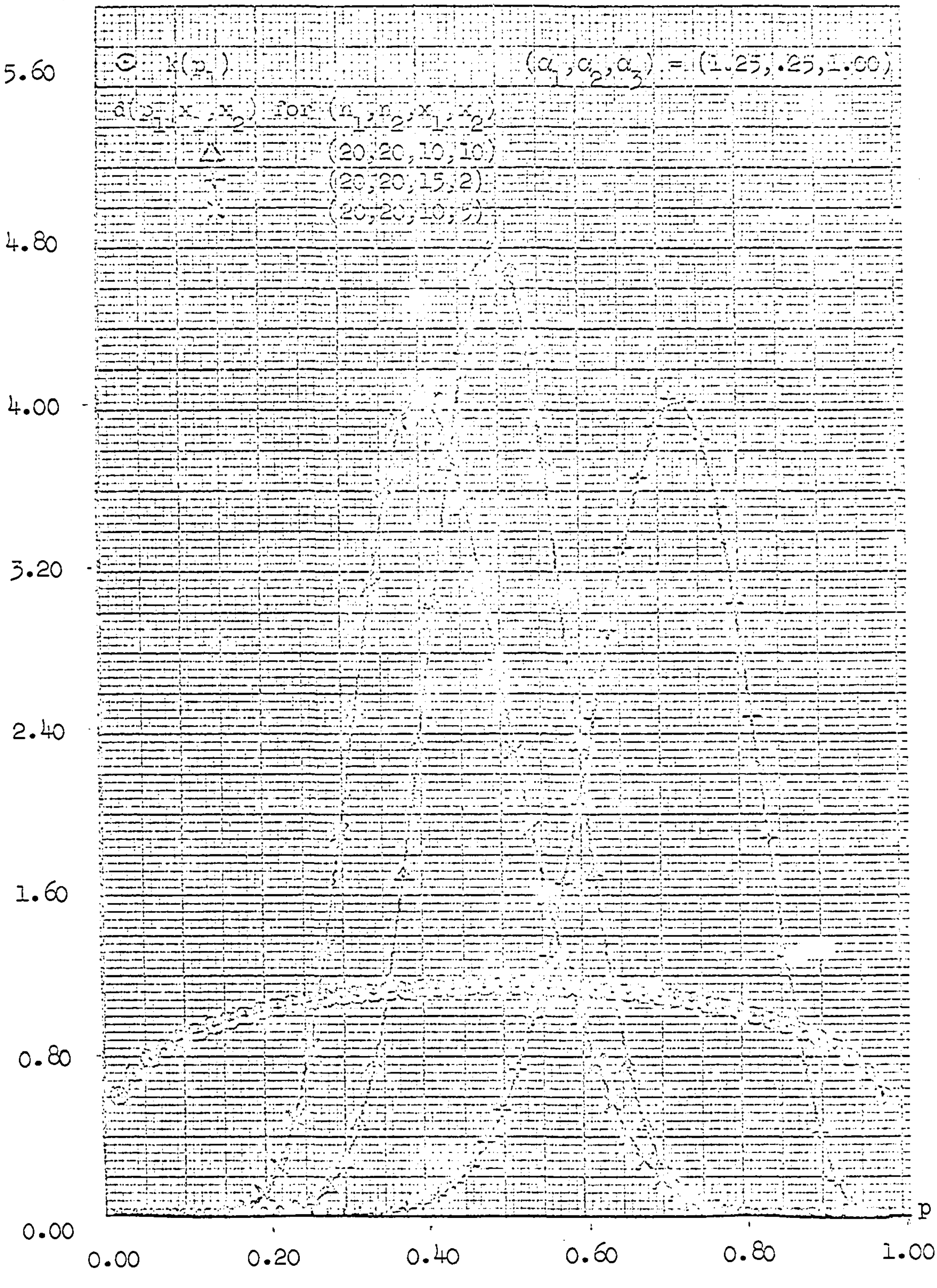

Figure 9.1. Prior and posterior distributions of $\tilde{p}_1$

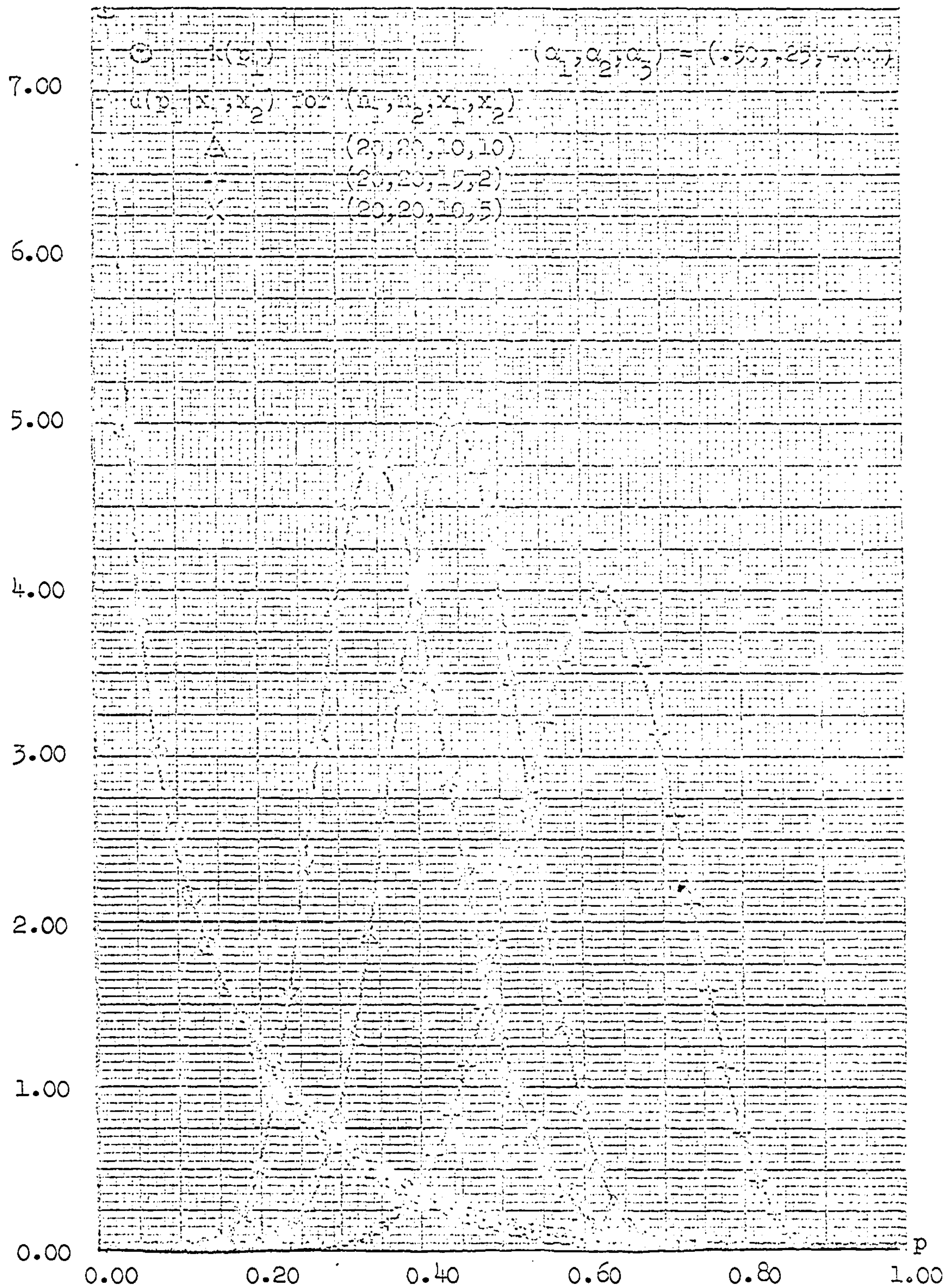

Figure 9.2. Prior and posterior distributions of $\tilde{p}_1$

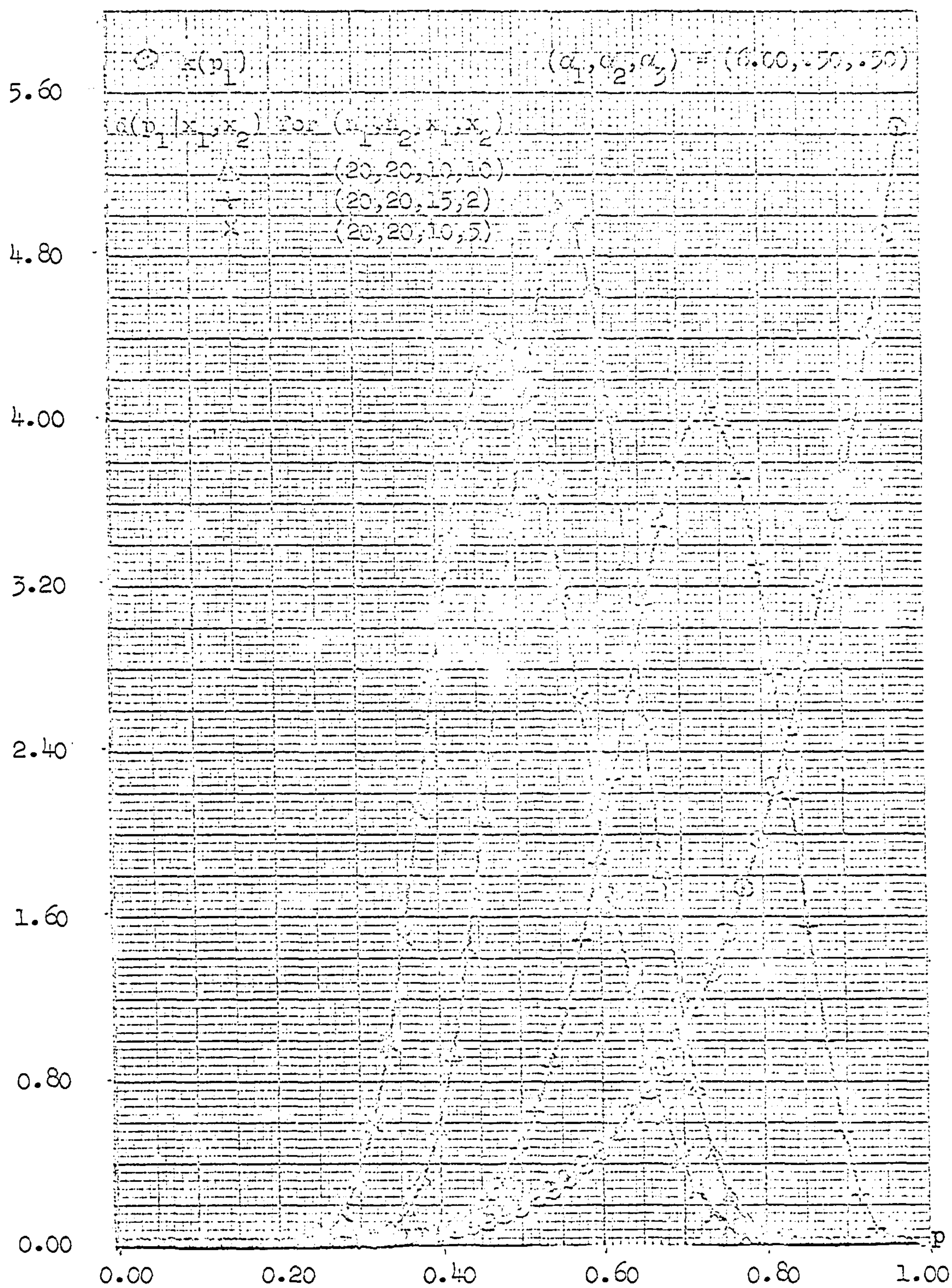

Figure 9.3. Prior and posterior distributions of $\tilde{p}_1$

X. LITERATURE CITED

Aitchison, J. 1964. Bayesian tolerance regions. Royal Statistical Society Journal Series B, 26: 161-175.

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York, New York, John Wiley and Sons, Inc.

Ando, Albert and Kaufman, G. M. 1965. Bayesian analysis of the independent multinormal process: neither mean nor precision known. American Statistical Association Journal 60: 347-358.

Bancroft, T. A. 1944. On biases in estimation due to the use of preliminary tests of significance. Annals of Mathematical Statistics 15: 190-204.

Bancroft, T. A. 1964. Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. Biometrics 20: 427-442.

Bartholomew, D. J. 1964. Discussion on the papers by Mr. Aitchison and Mr. Thatcher. Royal Statistical Society Journal Series B, 26: 192-210.

Bennett, B. M. 1952. Estimation of means on the basis of preliminary tests of significance. Institute of Statistical Mathematics Annals 4: 31-43.

Bennett, B. M. 1956. On the use of preliminary tests in certain statistical procedures. Institute of Statistical Mathematics Annals 8: 45-52.

Berkson, J. 1942. Tests of significance considered as evidence. American Statistical Association Journal 37: 325-335.

Cochran, W. G. 1963. Sampling techniques. New York, New York, John Wiley and Sons, Inc.

Cramèr, Harald. 1946. Mathematical methods of statistics. Princeton, New Jersey, Princeton University Press.

Edwards, W., Lindman, H., and Savage, L. J. 1963. Bayesian inference for psychological research. Psychological Review 70: 193-242.

Evans, T. G. 1965. Bayesian estimation of parameters of a multivariate normal distribution. Royal Statistical Society Journal Series B, 27: 279-283.

Fiering, Myron B. 1962. On the use of correlation to augment data. American Statistical Association Journal 57: 20-32.

Foster, F. G. and Rees, D. H. 1957. Upper percentage points of the generalized Beta distribution. I. Biometrika 44: 237-247.

Geisser, Seymour. 1965a. A Bayes approach for combining correlated estimates. American Statistical Association Journal 60: 602-607.

Geisser, Seymour. 1965b. Bayesian estimation in multivariate analysis. Annals of Mathematical Statistics 36: 150-159.

Geisser, Seymour and Cornfield, Jerome. 1963. Posterior distributions for multivariate normal parameters. Royal Statistical Society Journal Series B, 25: 368-376.

Ghosh, B. 1949. Interpenetrating networks of samples. Calcutta Statistical Association Bulletin 2: 108-119.

Good, Irving John. 1965. The estimation of probabilities (an essay on modern Bayesian methods). Cambridge, Massachusetts, The MIT Press.

Halperin, M. 1961. Almost linearly-optimum combination of unbiased estimates. American Statistical Association Journal 56: 36-43.

Huntsberger, D. V. 1955. A generalization of a preliminary testing procedure for pooling data. Annals of Mathematical Statistics 26: 734-743.

Jeffreys, H. 1961. Theory of probability. 3rd edition. London, England, Oxford University Press.

Kale, B. K. and Bancroft, T. A. 1967. Inference for some incompletely specified models involving normal approximations to discrete data. Biometrics 23: 335-348.

Khatri, C. G. and Pillai, K. C. S. 1965. Some results on the non-central multivariate Beta distribution and moments of traces of two matrices. Annals of Mathematical Statistics 36: 1511-1520.

Kitagawa, T. 1954. Empirical functions and interpenetrating sampling procedures. Kyūsyū University Memoirs of the Faculty of Science Series A, 8: 109-152.

Kitagawa, T. 1956. Some contributions to the design of sample surveys. Sankhyā 17: 1-36.

Kitagawa, T. 1963. Estimation after preliminary tests of significance. University of California Publications in Statistics 3: 147-186.

Krishnaiah, P. R. 1965. On the simultaneous ANOVA and MANOVA tests. Institute of Statistical Mathematics Annals 17: 35-53.

Krishnaiah, P. R. and Armitage, Lt. Col. J. V. 1965. Probability integrals of the multivariate F distribution, with tables and applications. Wright-Patterson Air Force Base, Ohio, Aerospace Research Laboratories, Office of Aerospace Research.

Kshirsagar, A. M. 1961. The non-central multivariate Beta distribution. Annals of Mathematical Statistics 32: 104-111.

Lindley, D. V. 1961. The use of prior probability distributions in statistical inference and decisions. Berkeley Symposium on Mathematical Statistics and Probability, 4th, 1960, Proceedings 1: 453-468.

Mauldon, J. G. 1959. A generalization of the Beta distribution. Annals of Mathematical Statistics 30: 509-520.

Mokashi, V. K. 1949. A note on interpenetrating samples. Indian Society of Agricultural Statistics Journal 2: 189-195.

Mosimann, James E. 1962. On the compound multinomial distribution, the multivariate Beta distribution, and correlations among proportions. Biometrika 49: 65-82.

Mosteller, F. 1948. On pooling data. American Statistical Association Journal 43: 231-242.

Olkin, Ingram and Rubin, Harold. 1964. Multivariate Beta distributions and independence properties of the Wishart distribution. Annals of Mathematical Statistics 35: 261-269.

Olkin, Ingram and Rubin, Harold. 1966. Correction to multivariate Beta distributions and independence properties of the Wishart distribution. Annals of Mathematical Statistics 37:297.

Raiffa, Howard and Schlaifer, Robert. 1961. Applied statistical decision theory. Boston, Massachusetts, Harvard Business School Division of Research.

Roberts, Harry V. 1966. Statistical dogma: one response to a challenge. American Statistician 20, No. 4: 25-27.

Savage, Leonard J. 1961a. The foundations of statistics reconsidered. Berkeley Symposium on Mathematical Statistics and Probability, 4th, 1960, Proceedings 1: 575-586.

Savage, Leonard J. 1961b. The subjective basis of statistical practice. Unpublished multilithed notes. Ann Arbor, Michigan, Department of Mathematics, University of Michigan.

Savage, Leonard J. 1962. Subjective probability and statistical practice. In Bartlett, M. S., ed. The foundations of statistical inference. pp. 9-35. New York, New York, John Wiley and Sons, Inc.

Smith, Cedric A. B. 1961. Consistency in statistical inference and decision. Royal Statistical Society Journal Series B, 23: 1-25.

Stone, M. 1963. The posterior t-distribution. Annals of Mathematical Statistics 34: 568-573.

Stone, M. 1964. Comments on a posterior distribution of Geisser and Cornfield. Royal Statistical Society Journal Series B, 26: 274-276.

Thatcher, A. R. 1964. Relationship between Bayesian and confidence limits for predictions. Royal Statistical Society Journal Series B, 26: 176-192.

Tikkiwal, B. D. 1960. On the theory of classical regression and double sampling. Royal Statistical Society Journal Series B, 22: 131-138.

Welch, B. L. 1964. Discussion on the papers by Mr. Aitchison and Mr. Thatcher. Royal Statistical Society Journal Series B, 26: 192-210.

Wilks, Samuel S. 1962. Mathematical statistics. New York, New York, John Wiley and Sons, Inc.

Zacks, S. 1966. Unbiased estimation of the common mean of two normal distributions based on small samples of equal size. American Statistical Association Journal 61: 467-476.

XI. ACKNOWLEDGEMENTS

I am indebted to the following persons and institutions for the successful completion of my graduate work at Iowa State University: to Dr. Joseph Sedransk, for suggestion of this dissertation topic, encouragement, and guidance; to The Iowa State Research Foundation for the financial support of my first year of graduate work at Iowa State; to The National Institutes of Health Biometry Training Grant 5T1GM34 for the financial support of my remaining years of graduate work at Iowa State; to Dr. T. A. Bancroft for encouragement; to Ronald and Carol Fuchs for hospitality; and to my husband, Charles, for patience.